Beyond CT scans: IMAGING IN 5D

Diverse genetic data helps PREDICT DISEASE

Peeling back layers of HIDDEN HISTORY

AMERICAN
Scientist
September–October 2023
www.americanscientist.org

A New Look At Exoplanets

Interacting light waves could boost the search for other Earths.



SIGMA XI
THE SCIENTIFIC RESEARCH HONOR SOCIETY



"I can't begin to describe this gorgeous ring...it is absolutely stunning. The workmanship and style is out of this world. Please order this ring, you won't be disappointed."
— J., Stuart, FL



Receive these scintillating Ultimate Diamond Alternative™, Diamond*Aura*® Classique sterling silver stud earrings **FREE!**

Read details below.

The Fifth C?

Cut, Color, Carat, Clarity…Chemistry?

Is it possible that the mind of a scientist can create more beauty and romance than Mother Nature? The Ultimate Diamond Alternative®, Diamond*Aura*®, was created with one mission in mind: Create brilliant cut jewelry that allows everyone to experience more clarity, more scintillation and larger carat weights than they have ever experienced. So, we've taken 2 ½ carats of our Ultimate Diamond Alternative® Diamond*Aura*® and set them in the most classic setting—the result is our most stunning, fiery, faceted design yet! In purely scientific measurement terms, the refractory index of these stones is very high, and the color dispersion is ***actually superior*** to mined diamonds.

Perfection from science. We named our brilliant cut diamond alternative stones Diamond*Aura* because "they dazzle just like natural diamonds but without the outrageous cost." We will not bore you with the incredible details of the scientific process, but will only say that it involves the use of rare minerals heated to an incredibly high temperature of nearly 5000°F. This can only be accomplished inside some very modern and expensive laboratory equipment. After several additional steps, scientists finally created a clear marvel that looks even better than the vast majority of mined diamonds. According to the book *Jewelry and Gems–the Buying Guide*, the technique used in our diamond alternative Diamond*Aura*® offers, "The best diamond simulation to date, and even some jewelers have mistaken these stones for mined diamonds."

COMPARE THE BRILLIANCE, FIRE & PRICE

	Mined Flawless Diamond	Diamond*Aura* Classique Ring
Hardness	Cuts Glass	**Cuts Glass**
Cut (58 facets)	Brilliant	**Brilliant**
Color	"D" Colorless	**"D" Colorless**
Clarity	"IF"	**Clear**
Dispersion/Fire	0.044	**0.066**
2½ ctw ring	$60,000+	**$39**

The 4 C's. Our 3-Stone Classique Ring retains every jeweler's specification: color, clarity, cut, and carat weight. The transparent color and clarity of our diamond alternative Diamond*Aura*® emulate the most perfect diamonds—D Flawless, and both are so hard they will cut glass. The brilliant cut maximizes the fire and radiance of the stone so that the light disperses into an exquisite rainbow of colors.

Rock solid guarantee. This .925 sterling silver ring is prong-set with a 1 ½ carat round brilliant stone in the center, showcased between two round brilliants stones of ½ carat each. Adding to your 4 C's, we will include our Ultimate Diamond Alternative® Diamond*Aura*® stud earrings for **FREE!** Try the Diamond*Aura*® 3-Stone Classique Ring at **$39** for 30 days. If for any reason you are not satisfied with your purchase, simply return it to us for a full refund of the item purchase price and keep the stud earrings as our gift.

Not Available in Stores

3-Stone Classique Ring
(2 ½ ctw) ~~$299†~~ **$39** + S&P Save $260

FREE stud earrings with purchase of Classique Ring—a $99 value!

Available in whole sizes 5-10

You must use the offer code to get our special price.

1-800-333-2045

Your Offer Code: DAR986-05

† Special price only for customers using the offer code versus the price on Stauer.com without your offer code.

Stauer®

14101 Southcross Drive W.,
Ste 155, Dept. DAR986-05,
Burnsville, Minnesota 55337
www.stauer.com

AMERICAN
Scientist

Volume 111 • Number 5 • September–October 2023

Departments

258 From the Editors

259 Online Preview

260 In Memoriam
Henry Petroski, 1942–2023

262 Spotlight
Hurricanes push heat into oceans • Changing policies on COVID-19 transmission • Briefings

272 Sightings
Pushing and pulling in plants

274 Perspective
Direct detection of exoplanets
Marija Strojnik

280 Technologue
Unearthing a prayer for the dead
David S. Citrin, Alexandre Locquet, and Junliang Dong

Scientists' Nightstand

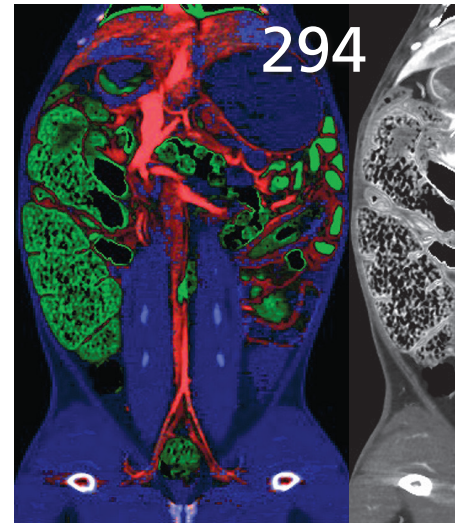
310 Book Reviews
An ocean of reasons to map the seafloor • Birdwatching as a social lens • Climate change is here: What now?

From Sigma Xi

317 Sigma Xi Today
Scientists as dream-sowers • IFoRE keynote speakers • Fall nomination and submission deadlines • STEM art and film at IFoRE • Faces of GIAR: Emma Dougherty • Spring GIAR awards • Regeneron ISEF award winners

Feature Articles





286 Genomic Prediction in the Big Data Era
A simple model from the early 20th century remains our best tool for using DNA to predict disease risk and other complex traits.
Gustavo de los Campos and Daniel Gianola

294 Medical Imaging in Increasing Dimensions
Combining technologies and moving into virtual space makes seeing into the body more accurate and beneficial.
Ge Wang

302 Time Is an Object
Not a backdrop, an illusion, or an emergent phenomenon, time has a physical size that can be measured in laboratories.
Sara Walker and Lee Cronin



The Cover



The search for planets outside of our Solar System, or exoplanets, has taken off since the first one was detected in the 1990s; now thousands have been identified (such as those in this artist's conception of the inner rocky planets of the TRAPPIST-1 system). However, all the methods currently used to find exoplanets have drawbacks, such as the requirement of a large time gap between measurements, which increases the potential for errors. In this issue's Perspective column, "Direct Detection of Exoplanets" (pages 274–279), optical scientist Marija Strojnik discusses a method adapted from optical testing that uses the wave nature of light, which allows measurements to be made over a short time, potentially increasing accuracy. This method takes advantage of the off-axis position of an exoplanet orbiting its star, which alters the tilt of light wave fronts coming from the planet in comparison with those coming from its star. (Cover artwork by NASA, ESA, CSA, Joseph Olmsted/STScI.)

Seeing What We Can't See



More than 5,000 *exoplanets*—planets that are outside our Solar System—have been identified so far. But we haven't actually seen any. So how have they been identified? Most often, detection techniques involve a disruption in another signal, such as a dip in the light from a star when its orbiting planet crosses in front of it, from the viewpoint of Earth. This phenomenon, called *transiting*, has been covered many times in *American Scientist*'s pages. The difficulty with exoplanet detection remains that most of the methods currently used have drawbacks, such as a lengthy time delay between measurements, making it more likely that errors in measurement will creep in. Optical scientist Marija Strojnik describes another potential detection method that uses the wave nature of light, which could allow for more accuracy. In "Direct Detection of Exoplanets" (pages 274–279), Strojnik shows that curved light wave fronts coming from distant exoplanet systems become planar over long distances, and those wave fronts create a detectable signal. The wave front from an exoplanet will develop a tilt in relation to the wave front from its star, because of the planet's off-axis position. Those two wave fronts can be combined by a detector, essentially creating an interference pattern that would not exist if the exoplanet wasn't there.

David S. Citrin, Alexandre Locquet, and Junliang Dong are similarly trying to uncover something that has been optically invisible, as they describe in this issue's Technologue column. In "Unearthing a Prayer for the Dead" (pages 280–285), Citrin, Locquet, and Dong tell us how they were able to see past corrosion so thick on a historic lead cross that it had completely obscured the inscription. The team used waves in a length between microwaves and infrared, in what is called the *terahertz band*, then used a series of computing techniques to isolate the inscription on the cross without causing any damage.

Preventing damage is also the focus of Ge Wang, whose goal is to combine a wide variety of medical imaging modes into a single, quick scan. In "Medical Imaging in Increasing Dimensions" (pages 294–301), Wang recounts early research in medical imaging, starting from x-rays in the 1890s, and how these methods have been combined and expanded in the modern era to create images in three dimensions, as well as the additional dimensions of color, time, and perhaps even augmented reality. Wang's hope is to combine computed tomography (CT) and magnetic resonance imaging (MRI) into one scan, which would greatly benefit patients for whom treatment time is of the essence, for example those who have just suffered from a stroke. Farther in the future, Wang posits that virtual avatars of patients could be made, which could be virtually scanned to determine which imaging technique would yield the best result for each situation, without exposing the patient to excess radiation.

We hope you enjoy this issue's full range of articles, from a feature on using genomes and big data to predict disease risk and other complex traits (pages 286–293), to another feature that delves deeply into the nature of time using a theory that links time to the influence of biological life—and which could help in the search for life beyond Earth (pages 302–309). Do you have feedback on any of our content? Readers are always welcome to send us an email, contact us through our website form, or post comments on social media. We look forward to hearing from you.

—Fenella Saunders (@FenellaSaunders)

AMERICAN
Scientist
www.americanscientist.org
VOLUME 111, NUMBER 5

EDITORIAL
Editor-in-Chief Fenella Saunders
Special Issue Editor Corey S. Powell
Managing Editor Stacey Lutkoski
Digital Features Editor Katie L. Burke
Acting Book Review Editor Jaime Herndon
Senior Contributing Editors Efraín E. Rivera-Serrano, Sarah Webb
Contributing Editors Sandra J. Ackerman, Emily Buehler, Amy Dusto, Christa Evans, Jeremy Hawkins, Karen Hudes, Flora Taylor
Editorial Associate Mia Evans

ART
Art Director Barbara J. Aulicino

DIGITAL
Acting Digital Managing Editor Kindra Thomas

ADMINISTRATION
EDITORIAL CORRESPONDENCE
American Scientist
P.O. Box 13975
Research Triangle Park, NC 27709
919-549-4691 • editors@amscionline.org

CIRCULATION AND MARKETING
NPS Media Group • Jamie Fallon, *account director*

ADVERTISING SALES
advertising@amsci.org • 800-243-6534

SUBSCRIPTION CUSTOMER SERVICE
American Scientist
P.O. Box 193
Congers, NY 10920
800-282-0444 • custservice@amsci.org

PUBLISHER
SIGMA XI, THE SCIENTIFIC RESEARCH HONOR SOCIETY
President Marija Strojnik
Treasurer David Baker
President-Elect Kathy Lu
Immediate Past President Nicholas A. Peppas
Executive Director & Publisher Jamie L. Vernon

EDITORIAL ADVISORY PANEL
Richard Boudreault, *University of Waterloo*
René Fuanta, *East Stroudsburg University*
Simson Garfinkel, *AI2050, Schmidt Futures*
Sonya T. Smith, *Howard University*
Caroline VanSickle, *A. T. Still University of Health Sciences*

American Scientist gratefully acknowledges support for engineering content through the Leroy Record Fund.

Sigma Xi, The Scientific Research Honor Society is a society of scientists and engineers, founded in 1886 to recognize scientific achievement. A diverse organization of members and chapters, the Society fosters interaction among science, technology, and society; encourages appreciation and support of original work in science and technology; and promotes ethics and excellence in scientific and engineering research.

Printed in the USA

Why We Podcast: Sharing Your Voice in Science
In episode five of *DEAI ComSci* (for Diversity, Equity, Accessibility, and Inclusion), *American Scientist*'s science-for-all podcast, host Jordan Anderson discusses how podcasts can help us navigate complex, relevant issues in science and society. Interviews include Sameer Honwad (professor at the University of Buffalo), Wendy Zukerman (host of the *Science Vs* podcast and 2022 Sigma Xi honorary membership award winner), and Aaron Scott and Emily Kwong from NPR's *Short Wave* podcast, as well as one of their guests, Dr. Tiana Williams-Claussen from the Yurok Tribe.
www.amsci.org/node/5106

Obesity's Lasting Effect on Brain Chemical Signals
When Mireille Serlie has clinic hours in her role as an endocrinologist, she sees obese patients who tell her that they know they have eaten a big meal, but they don't have any feeling of being full. Serlie has taken that observation into her research work on obesity. She and her colleagues have found that people with obesity appear to be correct: Their brains aren't giving them the proper signals in response to nutrients. And what's even more confounding is that when these patients lose significant weight, their brain responses do not return to normal. Learn more about this new study in this exclusively online blog article.
www.amsci.org/node/5109

The Cardiovascular Health Risks from the Stress of Racism
In this recent talk in *American Scientist*'s speaker series, Science by the Slice, Vanessa Volpe of North Carolina State University discusses her research group's findings on the effects of racial exclusion as a psychosocial stressor that can lead to long-term physiological changes such as hypertension and cardiovascular disease. She also looks at the role of anticipatory and intergenerational stress in health outcomes.
www.amsci.org/node/5110

Check out *AmSci Blogs*
www.amsci.org/blog/

Find *American Scientist* on Facebook
facebook.com/AmericanScientist

Follow us on Twitter
twitter.com/AmSciMag

Join us on LinkedIn
linkedin.com/company/american-scientist

Find us on Instagram
instagram.com/american_scientist/

American Scientist (ISSN 0003-0996) is published bimonthly by Sigma Xi, The Scientific Research Honor Society, P.O. Box 13975, Research Triangle Park, NC 27709 (919-549-0097). Newsstand single copy $5.95. Back issues $7.95 per copy for 1st class mailing. U.S. subscriptions: one year print or digital $30, print and digital $36. Canadian subscriptions: one year print $38, digital $30; other foreign subscriptions: one year print $46, digital $30. Print institutional rate: $75; Canadian $83; other foreign $91. Digital site license $200, print and digital institutional rate $275. Copyright © 2023 by Sigma Xi, The Scientific Research Honor Society, Inc. All rights reserved. No part of this publication may be reproduced by any mechanical, photographic, or electronic process, nor may it be stored in a retrieval system, transmitted, or otherwise copied, except for onetime noncommercial, personal use, without written permission of the publisher. Periodicals postage paid at Durham, NC, and additional mailing offices. Postmaster: Send change of address form 3579 to American Scientist, P.O. Box 193, Congers, NY 10920. Canadian publications mail agreement no. 40040263.

Travel Opportunities for October 2023 - May 2024



***Grand Staircase Utah Annular Solar Eclipse* - October 9 - 17, 2023**
Discover the North Rim of the Grand Canyon, Monument Valley, and Bear Ears National Monument. Then see the Annular Eclipse over Capitol Reef Nat'l Park and explore Grand Staircase Nat'l Monument and Bryce Canyon. With outstanding geologist and naturalist, Dr. Andre Potochnik.

Discover Egypt including Abydos & Dendara
November 4 - 16, 2023
Come explore the great historic and cultural sites of Egypt! From the pyramids and magnificent museums in Cairo to the Valley of the Kings near Luxor, and the majestic site of Abu Simbel in Upper Egypt. Travel on board a comfortable Nile Cruise vessel from Luxor to Aswan. With an excellent local archaeological guide.



Galapagos Islands
Dec 1 - 10, 2023 or Jan 26 - Feb 4, 2024
Discover Darwin's enchanted isles on board the National Geographic Endeavour II. Travel on board ship from island to island as you discover the endemic birdlife, marine iguanas, volcanic landscapes, and more!



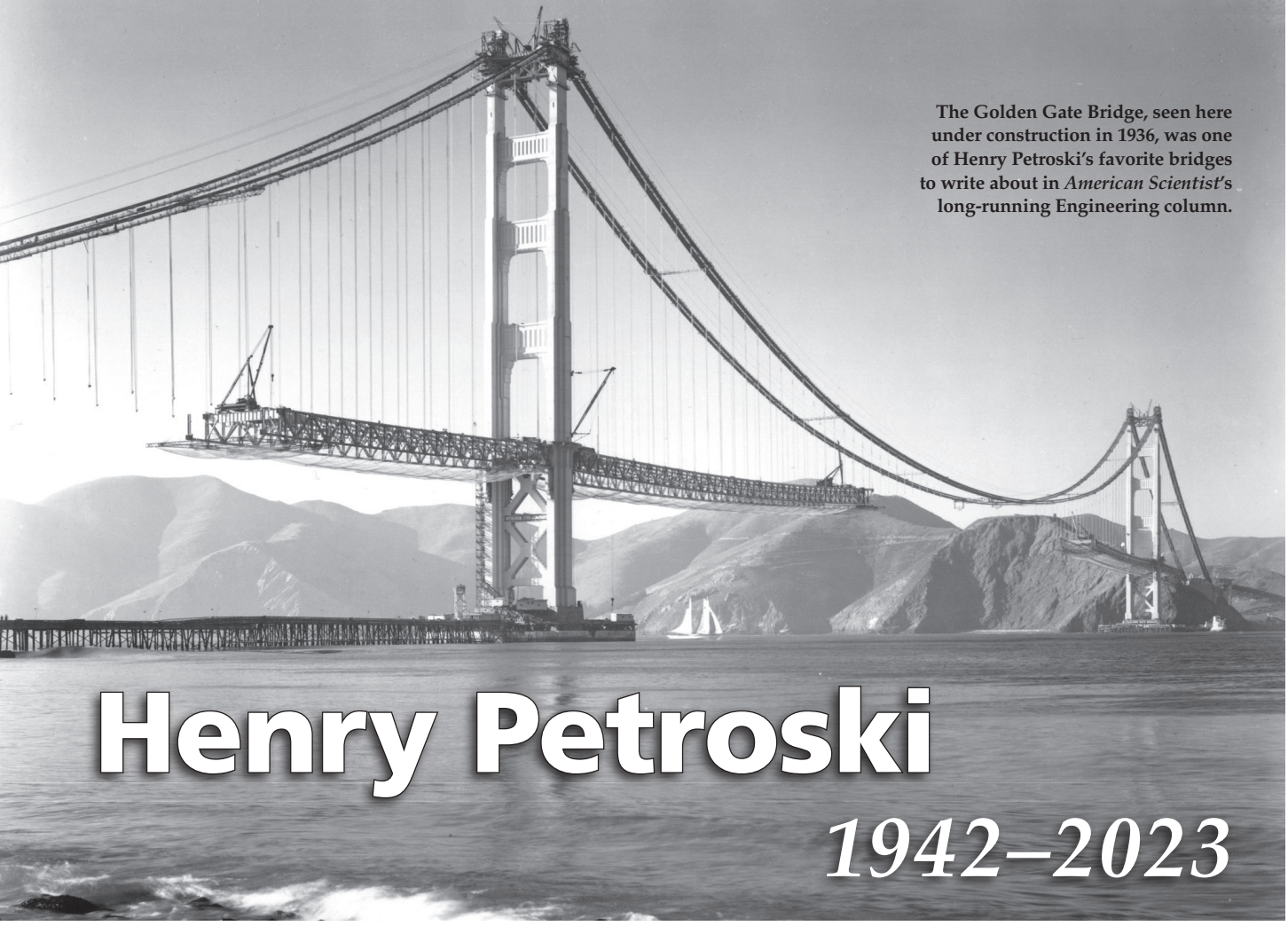
Alaska Aurora Borealis
March 7 - 13, 2024
Come see the Greatest Light Show on Earth and discover the wild beauty of Alaska in winter and the Aurora Borealis! From Anchorage and Talkeetna, take the train to Fairbanks, past Denali, the highest mountain in North America. See sled dogs, the Ice Festival, and the Aurora Borealis in the night skies.

***California Spring in Death Valley & Joshua Tree Nat'l Parks* - April 14 - 22, 2024**
Discover spring amidst the multi-colored landscapes in California's magnificent desert national parks. Learn about this geological wonderland, look for spring wildflower bloom, and enjoy the stunning vistas in balmy spring weather.

We invite you to travel the World with Sigma Xi!

SIGMA XI *Expeditions*
THE SCIENTIFIC RESEARCH HONOR SOCIETY
Phone: (800) 252-4910

For information please contact:
Betchart Expeditions Inc.
17050 Montebello Rd, Cupertino, CA 95014-5435
Email: SigmaXiInfo@betchartexpeditions.com



The Golden Gate Bridge, seen here under construction in 1936, was one of Henry Petroski's favorite bridges to write about in *American Scientist*'s long-running Engineering column.

Henry Petroski

1942–2023

Golden Gate Bridge, Highway and Transportation

Our esteemed colleague and the longtime author of the Engineering column, Henry Petroski, passed away on June 14, 2023, at 81 years old.

Henry wrote almost 200 articles for the Engineering column, having authored it in every issue since January–February 1991. I've had the honor of editing Henry's columns since 2013. My predecessor, David Schoonmaker, summed up well what editing the column was like: "Working with Henry Petroski was a highlight of my time at *American Scientist*. The obvious delight he took in developing a story to make complex engineering understandable was infectious. And by creating such a narrative, Henry was able to reveal engineering as both science and humanity."

Henry was a dedicated writer—he turned in what was to be his final column for *American Scientist* just a few months before his death, and he passed away shortly after it went to press. His 20th book, *Force: What It Means to Push and Pull, Slip and Grip, Start and Stop*, came out in 2022. He epitomized the prolific career, to a level that most of us can only aspire to.

Henry's final column for *American Scientist*, in our special single-topic issue on scientific modeling, covered some of his favorite subjects, including bridges, failure, and the importance of incorporating real-world mechanics into design. As he pointed out in that column, a brilliant idea is all very well and good, however: "That idea may conceivably promise to be the greatest thing since sliced bread, but it may have to forever remain at best a half-baked, private thought if it cannot be communicated to someone who can engineer it into something concrete." The column examined why small models often cannot be linearly scaled up to the largest structures, because mechanics can change with dimension. To explain this idea, Henry used one of his favorite examples of engineering failure, the Tacoma Narrows Bridge, which famously shook itself apart under high winds in 1940.

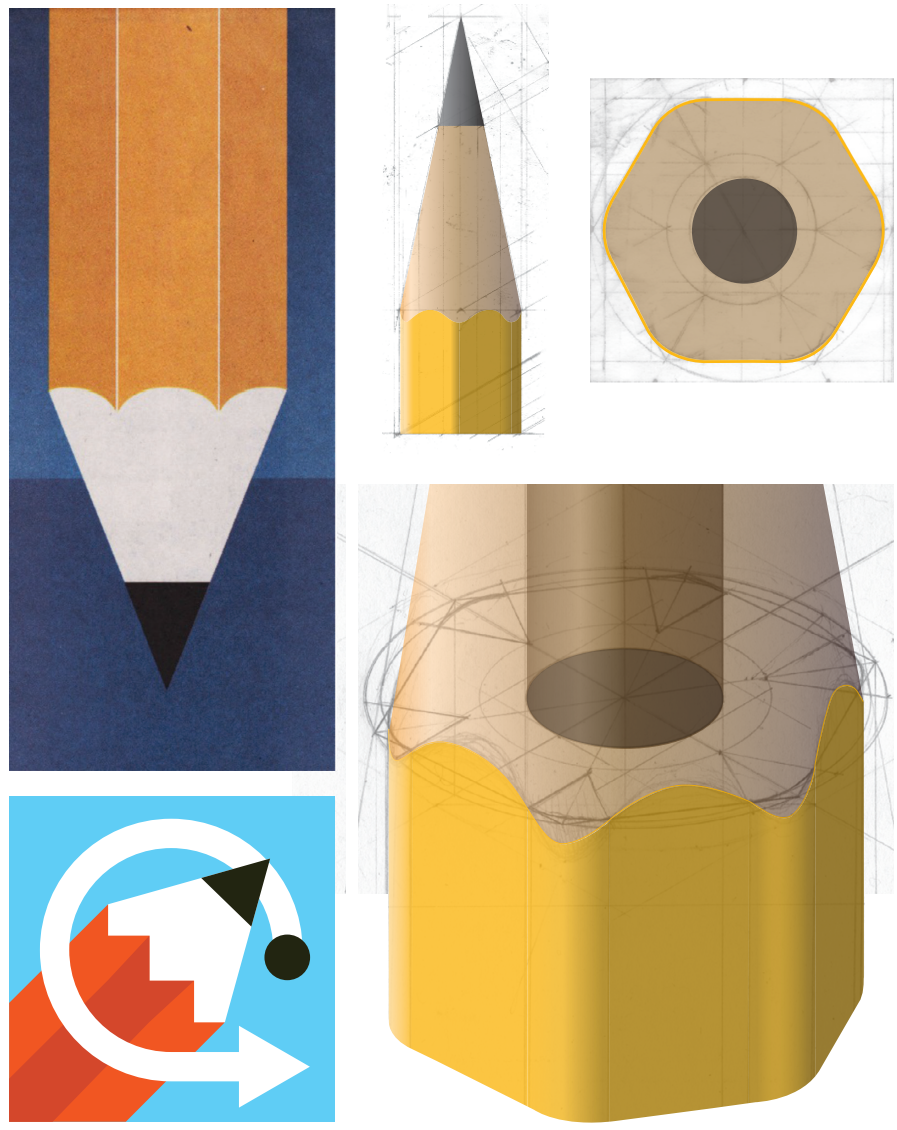
In Henry's May–June 2023 column, "Museums of Bridges," he enthusiastically included himself among the world's *pontists*, taken after the French word for bridge, a moniker adopted by historical bridge devotees. Regular readers of Henry's column would not be surprised by his declared love for bridges. Over the years, Henry's columns featured everything from the most famous bridges, such as the Brooklyn Bridge (most recently in January–February 2022) and the Golden Gate Bridge (most recently in January–February 2023), to the less well-known, such as the Genoa Bridge in Italy (September–October 2020) and Government Bridge across the Mississippi River between Iowa and Illinois (July–August 2012). Henry had a way of imbuing bridges with an almost human personality, describing what made each span unique, detailing its often complex history, and describing its importance to a region. He would draw readers into his topic and make them consider objects that they might not have otherwise noticed.

Henry was also well-known for his deep examinations of what he had himself originally thought of as "seemingly frivolous subjects," everyday objects that we take for granted, but which upon examination have deeper history and wider lessons to divulge about mechanics. His most famous example was the pencil, the subject of

one of his early books. In his March–April 2000 column, Henry explained how he had come across the pencil as a subject of interest, spurred by his exposure to low-quality pencils. He noted that "it is not regularity but anomaly that attracts attention, and I thought increasingly of the pencil as representative of all manufactured items and a metaphor for engineering itself." Henry had a knack for digging into the origins of these often overlooked objects, exploring colorful historical anecdotes, warring patent holders, and stark cultural differences in how they evolved over time in different places. He revisited the subject of pencils in his March–April 2014 column as an example of another of his favored topics: how the misrepresentation of objects can affect our perception of how mechanics works. In that article, Henry lamented how illustrations of pencils often incorrectly show the direction of ridge lines on sharpened pencil points (*see figure at right*) and expanded on the need for design to match reality. As he put it: "Engineers rely a great deal on computer drafting to convey their ideas to machinists and manufacturers, but without a keen perception of what looks right and what wrong, an engineer may send a perfectly neat drawing of a totally inappropriate part to a totally confused shop."

In other columns, Henry continued to explore such everyday objects as the toothpick (November–December 2007), bag ties (May–June 2008), the many ways newspapers have been folded for thrown delivery (May–June 2002), and "pizza savers," the small plastic devices that stop a pie from being squashed during delivery (July–August 2011). My personal favorite, however, may have been his column on the pocket protector, which appeared in May–June 2014. This seemingly simple plastic device has a long history that Henry unveiled with his exploration of patent diagrams, which he followed up with a tie-in to the cultural significance of engineering during the space race. This column took off on social media with the hashtag #nerdpride, a label I can get behind, as could Henry and a lot of his readers.

Although Henry often employed detailed examinations of small objects to deliver bigger messages about engineering, he also often discussed a myriad of other topics, from skyscrapers to geothermal energy. And he did not shy away from commenting on larger systemic problems. His columns covered such hot-button topics as the persistent problem of misuse of government contracts for private gain, or *graft*, in infrastructure projects, which he called a "plague" on the nation (September–October 2016), as well as the problematic lack of professional oversight of the condition of American co-op and condominium buildings that could have been behind the 2021 deadly Champlain Towers South collapse in Florida (September–October 2021). Henry broke one of his own rules and wrote the latter column before the final official report on the collapse was released—an endeavor that can take years—because we had received so many reader requests for his analysis of the disaster.



Tom Dunne

As longtime *American Scientist* contributor Henry Petroski pointed out in his March–April 2014 column, the ridges of sharpened pencils are often illustrated incorrectly (*left, top and bottom*). A correctly rendered pencil (*right*) shows how the scalloped border dips rather than rises at its ridge lines.

Henry's shoes simply can't be filled: He was the first columnist for the Engineering column, and he will remain the only columnist for it. We have decided to retire the Engineering column in his honor. All forms of engineering will still be covered in the Technologue column and throughout the magazine. But Henry's passing leaves a permanent gap. He will be greatly missed.

Readers have always had a lot to say about Henry's Engineering column. If you would like to share a remembrance about Henry or about your favorite of his columns, please email us at editors@amscionline.org. Henry was also a longtime member and an inaugural fellow of Sigma Xi, and there is an obituary for him on the Sigma Xi website. —*Fenella Saunders*

Hurricanes Push Heat into Oceans

High winds at sea can disturb ocean layers, driving warm water into deep ocean currents that traverse the planet.

When a hurricane hits land, the destruction can be visible for years or even decades. Less obvious, but also powerful, is the effect hurricanes have on oceans.

In a new study, we show through real-time measurements that hurricanes don't just churn water at the surface. They can also push heat deep into the ocean in ways that can lock it up for years and ultimately affect regions far from the storm.

Heat is the key component of this story. It has long been known that hurricanes gain their energy from warm sea surface temperatures. This heat helps moist air near the ocean surface rise into the sky like a hot air balloon and form clouds taller than Mount Everest, which is why hurricanes generally form in tropical regions.

What we discovered is that hurricanes ultimately help warm the ocean, too, by enhancing the ocean's ability to absorb and store heat. And that can have far-reaching consequences.

When hurricanes mix heat into the ocean, that heat doesn't just resurface in the same place. We showed how underwater waves produced by the storm can push heat roughly four times deeper than mixing alone, sending it to a depth where the heat is trapped far from the surface. From there, deep sea currents can transport the heat thousands of kilometers. (*See "Operational Oceanography," March–April 2022.*) A hurricane that travels across the western Pacific Ocean and hits the Philippines could end up supplying warm water that heats the coast of Ecuador years later.

At Sea, Looking for Typhoons

For two months in the fall of 2018, we lived aboard the research vessel *Thomas G. Thompson* to record how the Philippine Sea responded to changing weather patterns. As ocean scientists, we study turbulent mixing in the ocean, and hurricanes and other tropical storms that generate this turbulence.

Skies were clear and winds were calm during the first half of our experiment. But in the second half, three major typhoons—as hurricanes are known in this part of the world—stirred up the ocean.

That shift allowed us to directly compare the ocean's motions with and without the influence of storms. In particular, we were interested in learning how turbulence below the ocean surface was helping transfer heat down into the deep ocean.

We measure ocean turbulence with an instrument called a *microstructure profiler*, which free-falls nearly 300 meters and uses a probe similar to a phonograph needle to measure the motions of water.

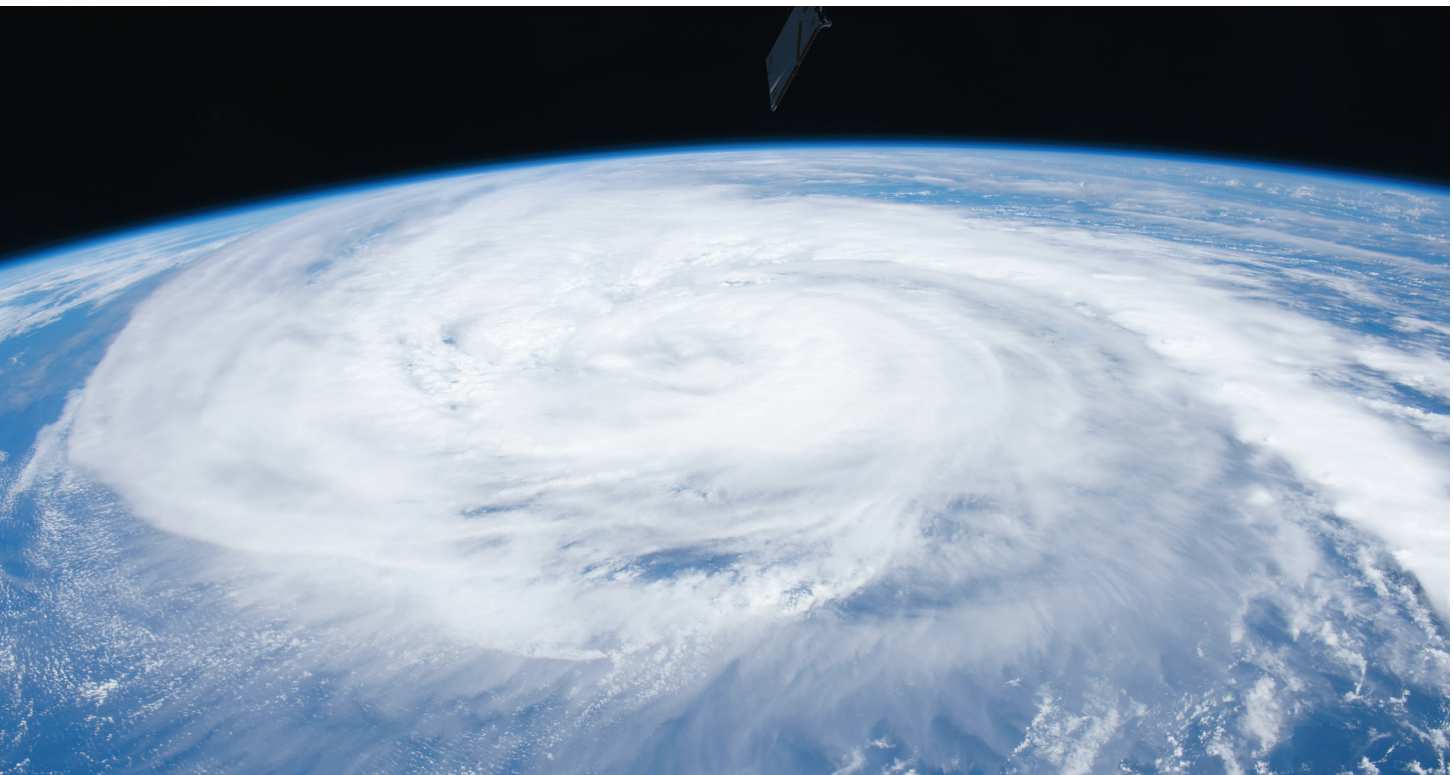
When a Hurricane Comes Through

Imagine the tropical ocean before a hurricane passes over it. At the surface is a layer of warm water, warmer than 27 degrees Celsius, that is heated by the Sun and extends roughly 50 meters below the surface. Below it are layers of colder water.

The temperature difference between the layers keeps the waters separated

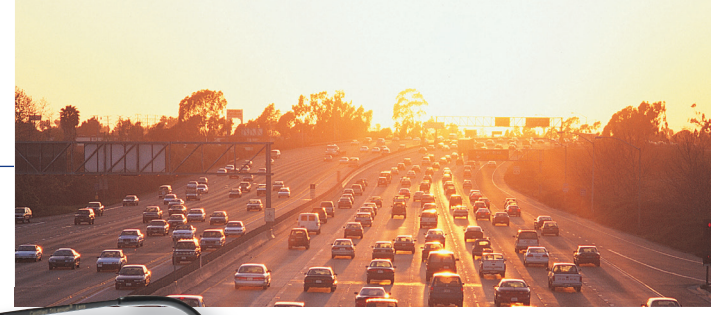
(continued on page 264)

Hurricanes—such as this 2021 typhoon in the Philippine Sea photographed from the International Space Station—disturb more than just the ocean surface. The high winds can also disturb the ocean's distinct layers, driving warm water deeper into the sea and pushing cold water upward. The ocean then holds onto that heat for months after the storm has passed.



NASA/Mark Garcia

Urgent: Special Summer Driving Notice



To some, sunglasses are a fashion accessory…

But When Driving, These Sunglasses May Save Your Life!

Drivers' Alert: Driving can expose you to more dangerous glare than any sunny day at the beach can… do you know how to protect yourself?

Studies by the National Highway Traffic Safety Administration (NHTSA) show that most (74%) of the crashes occurred on clear, sunny days



Navigator™ Black Stainless Steel Sunglasses

The sun rises and sets at peak travel periods, during the early morning and afternoon rush hours and many drivers find themselves temporarily blinded while driving directly into the glare of the sun. Deadly accidents are regularly caused by such blinding glare with danger arising from reflected light off another vehicle, the pavement, or even from waxed and oily windshields that can make matters worse. Early morning dew can exacerbate this situation. Yet, motorists struggle on despite being blinded by the sun's glare that can cause countless accidents every year.

Not all sunglasses are created equal. Protecting your eyes is serious business. With all the fancy fashion frames out there it can be easy to overlook what really matters—the lenses. So we did our research and looked to the very best in optic innovation and technology.

Sometimes it does take a rocket scientist. A NASA rocket scientist. Some ordinary sunglasses can obscure your vision by exposing your eyes to harmful UV rays, blue light, and reflective glare. They can also darken useful vision-enhancing light. But now, independent research conducted by scientists from NASA's Jet Propulsion Laboratory has brought forth ground-breaking technology to help protect human eyesight from the harmful effects of solar radiation light. This superior lens technology was first discovered when NASA scientists looked to nature for a means to superior eye protection—specifically, by studying the eyes of eagles, known for their extreme visual acuity. This discovery resulted in what is now known as Eagle Eyes®.



Slip on a pair of Eagle Eyes® and everything instantly appears more vivid and sharp. You'll immediately notice that your eyes are more comfortable and relaxed and you'll feel no need to squint. The scientifically designed sunglasses are not just fashion accessories—they are necessary to protect your eyes from those harmful rays produced by the sun during peak driving times.

The Only Sunglass Technology Certified by the Space Foundation for UV and Blue-Light Eye Protection. Eagle Eyes® features the most advanced eye protection technology ever created. The TriLenium® Lens Technology offers triple-filter polarization to block 99.9% UVA and UVB—plus the added benefit of blue-light eye protection. Eagle Eyes® is the only optic technology that has earned official recognition from the Space Certification Program for this remarkable technology. Now, that's proven science-based protection.

Certified EAGLE EYES® was developed from original NASA Optic technology and was recently inducted into the Space Foundation Technology Hall of Fame.

The finest optics: And buy one, get one FREE! Eagle Eyes® has the highest customer satisfaction of any item in our 20 year history. We are so excited for you to try the Eagle Eyes® breakthrough technology that we will give you a **second pair of Eagle Eyes® Navigator™ Sunglasses FREE—a $59.95 value!**

That's two pairs to protect your eyes with the best technology available for less than the price of one pair of traditional sunglasses. You get a pair of Navigators with stainless steel black frames and the other with stainless steel gold, plus one hard zipper case and one micro-fiber drawstring cleaning pouch are included. Keep one pair in your pocket and one in your car.

Your satisfaction is 100% guaranteed. If you are not astounded with the Eagle Eyes® technology, enjoying clearer, sharper and more glare-free vision, simply return one pair within 30 days for a full refund of the purchase price. The other pair is yours to keep. No one else has such confidence in their optic technology. Don't leave your eyes in the hands of fashion designers, entrust them to the best scientific minds on earth. Wear your Eagle Eyes® **Navigators** with absolute confidence, knowing your eyes are protected with technology that was born in space for the human race.

Receive the Navigator™ Gold Sunglasses (*a $59.95 value*) **FREE! just for trying the Navigator™ Black**



Navigator™ Gold Stainless Steel Sunglasses

Fit-ons available for $39 +S&H
Black or Tortoise-Shell design



Two Pairs of Eagle Eyes® Navigator™ Sunglasses ~~$119.90~~†

Offer Code Price **$49** + S&P **Save $70.90**

Offer includes one pair each Navigator™ Black and Navigator™ Gold Sunglasses

1-800-333-2045

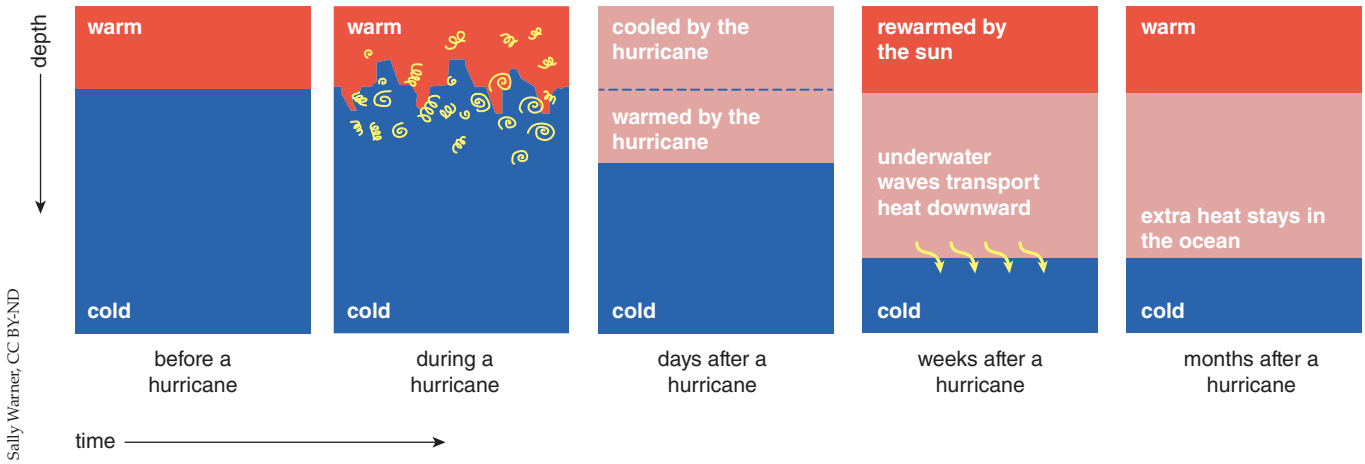
Your Insider Offer Code: EEN987-06
You must use this insider offer code to get our special price.

Stauer®
BBB ACCREDITED BUSINESS
Rating of A+

14101 Southcross Drive W., Ste 155, Dept. EEN987-06
Burnsville, Minnesota 55337
www.stauer.com

† Special price only for customers using the offer code versus the price on Stauer.com without your offer code.

Smart Luxuries—Surprising Prices™



Although the swapping of cold and warm ocean water might seem benign, the exchange has far-reaching consequences. Colder surface water more readily accepts heat from the atmosphere, and deeper warm water joins ocean currents, which can carry the heat for thousands of miles. The combined effect is an overall increase in ocean temperature.

(continued from page 262)
and virtually unable to affect each other. You can think of it like the division between the oil and vinegar in an unshaken bottle of salad dressing.

As a hurricane passes over the tropical ocean, its strong winds help stir the boundaries between the water layers, much like someone shaking the bottle of salad dressing. In the process, cold deep water is mixed upward from below and warm surface water is mixed downward. This process causes surface temperatures to cool, allowing the ocean to absorb heat more efficiently than usual in the days after a hurricane.

For more than two decades, scientists have debated whether the warm waters that are mixed downward by hurricanes could heat ocean currents and thereby shape global climate patterns. At the heart of this question was whether hurricanes could pump heat deep enough so that it stays in the ocean for years.

By analyzing subsurface ocean measurements taken before and after three hurricanes, we found that underwater waves transport heat roughly four times deeper into the ocean than direct mixing during the hurricane. These waves, which are generated by the hurricane itself, transport the heat deep enough that it cannot be easily released back into the atmosphere.

Heat in the Deep Ocean

Once this heat is picked up by large-scale ocean currents, it can be transported to distant parts of the ocean. The heat injected by the typhoons we studied in the Philippine Sea may have flowed to the coasts of Ecuador or California, following current patterns that carry water from west to east across the equatorial Pacific. At that point, the heat may be mixed back up to the surface by a combination of shoaling currents, upwelling, and turbulent mixing. Once the heat is close to the surface again, it can warm the local climate and affect ecosystems.

Underwater waves transport heat roughly four times deeper into the ocean than direct mixing during the hurricane.

For instance, coral reefs are particularly sensitive to extended periods of heat stress. El Niño events are the typical culprit behind coral bleaching in Ecuador, but the excess heat from the hurricanes that we observed may contribute to stressed reefs and bleached coral far from where the storms occurred.

It is also possible that the excess heat from hurricanes stays within the ocean for decades or more without returning to the surface, which would actually have a mitigating impact on climate change. As hurricanes redistribute heat from the ocean surface to greater depths, they can help to slow down warming of the Earth's atmosphere by keeping the heat sequestered in the ocean.

Scientists have long thought of hurricanes as extreme events fueled by ocean heat and shaped by the Earth's climate. Our findings, published in the *Proceedings of the National Academy of Sciences of the U.S.A.*, add a new dimension to this problem by showing that the interactions go both ways—hurricanes themselves have the ability to heat up the ocean and shape the Earth's climate. —*Noel Gutiérrez Brizuela and Sally Warner*

References

Glynn, P. W. 2003. Coral communities and coral reefs of Ecuador. In *Latin American Coral Reefs*, ed. Jorge Cortés, pp. 449–472. Amsterdam: Elsevier Science.

Gutiérrez Brizuela, N., et al. 2023. Prolonged thermocline warming by near-inertial internal waves in the wakes of tropical cyclones. *Proceedings of the National Academy of Sciences of the U.S.A.* 120:e2301664120.

Izumo, T. 2005. The equatorial undercurrent, meridional overturning circulation, and their roles in mass and heat exchanges during El Niño events in the tropical Pacific Ocean. *Ocean Dynamics* 55:110–123.

Moum, J. N., A. Perlin, J. D. Nash, and M. J. McPhaden. 2013. Seasonal sea surface cooling in the equatorial Pacific cold tongue controlled by ocean mixing. *Nature* 500:64–67.

Noel Gutiérrez Brizuela is a PhD student at the University of California, San Diego, with a background in physics and research focused on internal waves and ocean mixing, tropical oceanography, and nonlinear dynamical systems. Sally Warner is an associate professor of climate science at Brandeis University who studies how water moves and mixes in the ocean. This article is adapted from one that appeared in The Conversation, *theconversation.com. Email for Gutiérrez Brizuela: nogutier@ucsd.edu*

IFoRE '23
POWERED BY SIGMA XI

Student Research Awards

Present your research and have the opportunity to win awards!

Abstract deadline: September 1, 2023

Travel grants available

experienceIFoRE.org





Attention Members:

Look for your annual appeal letter in the mail after Labor Day to make your annual donation to Sigma Xi.

For first-time donors, use this QR code or email CJohnson@sigmaxi.org to support science and research.

sigmaxi.org/give

SIGMA XI
THE SCIENTIFIC RESEARCH HONOR SOCIETY

Changing Policies on COVID-19 Transmission



Photo by Peter Means courtesy of Virginia Tech

Despite its top-notch scientific institutions, the United States fared especially poorly during the COVID-19 pandemic. There were many missed opportunities that led to such an epic tragedy. One that has loomed especially large has been confusion around airborne spread of the virus. Precautions such as improving indoor air quality or wearing masks were ignored or downplayed until far too late. Linsey Marr, an engineer who studies aerosols at Virginia Tech, suddenly found her expertise needed in 2020 in the rapidly unfolding public health emergency—and yet she and others in her field were not included among the regulatory scientists working on the guidelines for the World Health Organization and U.S. Centers for Disease Control and Prevention (CDC). Digital features editor Katie L. Burke spoke with Marr about how she and her colleagues worked for years to change policies based on faulty ideas about transmission of the coronavirus. This interview has been edited for length and clarity.

How did you start working on airborne transmission?

Before 2009, I'd mainly been studying outdoor air pollution, particle emissions from vehicles. I became interested in airborne transmission when my son started day care in 2009. As with all kids who start day care, he got sick a lot. The day care center had what we consider good hygiene practices.

I started reading about how flus and colds spread. I was surprised to learn that at a mechanistic level, the way I think about particle pollution spreading through the air, we didn't actually know how viruses were getting from one person to another. There was the idea that they were coming out in these large, wet droplets that people would cough or sneeze onto each other. But frankly there wasn't a lot of evidence for that. They were missing out on the fact that people also release much tinier droplets, aerosol particles you can't see, that could potentially contain a virus. The first thing we did was collect particles from the air in a day care center, in the university health center, and on airplanes. We knew what size they were. In half our samples, we found the flu virus in these very tiny particles—not near a sick person, but just floating around in the air.

When the pandemic hit, what were you thinking about as you were hearing the first news about the virus?

In the outbreak of SARS 1 in 2003 and 2004, it had been pretty well established that that virus was transmitted through an airborne route. When I first heard about severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), my first assumption was, well, there's no reason why this wouldn't be too. Then, I was seeing the rapid spread in China, and then the locking down. I was seeing photos of health care workers in China in the hospitals wearing full protective gear. I knew you wouldn't go to the trouble of wearing that unless the virus was very transmissible. Those things tipped me off that SARS-CoV-2 was probably airborne too. In January 2020, I was thinking, yeah, this virus is probably airborne.

At that time, I also knew that the traditional thinking in medical and public health circles was that this virus would be transmitted by large droplets [as opposed to airborne aerosols] when people are close to each other, coughing on each other within six feet. There was a *New York Times* article published at the end of January, about six things that are important for understanding the spread of this virus. One of the questions was, How far does the virus travel? This article said this one doesn't travel very far, unlike measles and some others that can travel hundreds of feet or more. I thought that was totally wrong. I replied to the *New York Times* in a tweet and said, "I don't believe this. Why do people think this?"

Seeing things evolve throughout February 2020, it was clear that the U.S. medical and public health communities were doubling down on this idea that the virus was spreading just through these large droplets—over a distance no greater than six feet—and that we needed to worry about contaminated surfaces. I had my doubts. What clinched it for me was the Skagit Valley Chorale outbreak in early March 2020 in Washington State.

What was it about that outbreak that confirmed your suspicions that the coronavirus could be airborne?

It was the fact that the people going into that practice, the rehearsal, knew that this virus was there. They took some precautions. They avoided shaking hands. They spaced themselves out. Nevertheless, more than 50 people became infected at that one rehearsal. Even if they had shared food, or let's say everyone had touched the same doorknob, the idea that that doorknob or whatever object could become contaminated by one person, and then you could have 50 people come by and subsequently touch that same object, and there's still virus on it even after the first few people touch it? I just found that unbelievable.

Can you clarify the technical meaning of the terms *droplet*, *aerosol*, and *airborne*?

In the traditional definitions of disease transmission, the word *droplet* refers to things that are larger than 5 or 10 microns. These are assumed to travel kind of ballistically, following a certain trajectory as they come out of one person's mouth or nose, fly through the air, and land in someone else's eyes or nose or mouth, or land on a surface and contaminate it. That was thought to be the major way that colds and flu were transmitted. Then there are also *aerosols*, defined to be smaller than that cutoff. Those were thought to transmit mainly at long distances. Diseases such as measles and tuberculosis were

acknowledged to be transmitted by aerosols, to be *airborne*.

But the big blind spot medical researchers had in those definitions is that when you have these very small particles, they can transmit across a long distance or a close distance. When people are close to one another, they're even more likely to be exposed to these small particles.

Why was this idea that colds, flus, and SARS-CoV-2 spread by droplets rather than aerosols so entrenched? What was keeping people from acknowledging aerosols?

I don't know. I think a lot of the ideas about disease transmission were observations of who gets sick. After an outbreak, researchers would look for so-called close contacts, people who were in close proximity to the infected person. Then, they would find that these close contacts became sick. The thing is, because of the way their contact tracing was defined, they didn't really look at other people who were maybe in the room, but not closely associated with the infected person. They weren't looking for it, and so they didn't see it.

Another potential reason is that a lot of understanding of infectious disease transmission grew out of what happens in hospitals. Modern hospitals are pretty well ventilated. Under those conditions, you're less likely to see airborne transmission. If people are wearing surgical masks, they wouldn't completely block this route of transmission, but would help reduce it. And researchers and regulatory scientists could see these large, wet, visible droplets. It's easy to understand that those could carry and transmit virus. But for every one of those, there are hundreds of these tiny particles that you can't see. These scientists didn't know those smaller particles were there.

Medical researchers and MDs are experts on what's happening when a pathogen is inside your body. I have no expertise in that. Once it gets into your body, I don't know what happens. But if you're thinking about what happens to that pathogen when it's outside your body in the indoor or outdoor environment, I would argue that an environmental engineer or aerosol scientist has the right expertise to understand what is happening with that pathogen when it's moving around among people.

It seems like dividing transmission into this binary [droplet versus aerosol] doesn't make sense. I understand that from a policy perspective one often has to create categories and thresholds. But in this case, it ended up being confusing.

The way that the infection prevention and control community defined transmission was very binary. A disease was either transmitted by droplets or it was airborne. But in reality, when people are in close proximity, transmission could be via either route, large droplets or aerosol particles.

Before the pandemic, were you already working to change policies around airborne transmission? Or did that only happen once the pandemic arrived?

I had brought it up a few times. Other people who have expertise in this area also had brought it up but were ignored. I would say we were ridiculed, even, in some cases. We're outsiders to the field. The thought was, "Oh, what do these people know? They have no idea that this is how it's been defined for 50 years." It was kind of like shouting into the void. Obviously, this topic became much more critical during the pandemic, because if we continued to focus on droplets—staying six feet away from one another and wiping down surfaces—I didn't think we were going to control the disease very well.

In January 2020, I was thinking, yeah, this virus is probably airborne.

Lidia Morawska in Australia is also an expert in this field, and she pulled together a group of researchers worldwide who have expertise in how pathogens move around in the environment—most of us not MDs. She arranged a meeting with the WHO in early April 2020. I don't think they wanted to hear what we had to say. So, Morawska led efforts to publish a public letter to the WHO, saying, "We're pretty sure this is what's happening. You need to update your science so that we can control the spread better." We had 239 cosigners on this letter, which was published in July 2020 in *Clinical Infectious Diseases*.

It kind of took the WHO by surprise. They responded by acknowledging that airborne transmission can happen in special circumstances—for example, when there are what they called "aerosol-generating medical procedures" taking place in health care settings, such as intubating someone. They just were not aware that when people breathe and talk, they generate aerosols—actually more than are generated through these medical procedures.

Walk me through what happened over the course of the next couple of years, until policies really did start changing.

In August of 2020, there was a National Academies of Science, Engineering, and Medicine workshop on airborne transmission of SARS-CoV-2. Various experts spoke at that workshop and presented evidence for airborne transmission. It was a very strong case at that point. Coming out of that workshop, Kim Prather, who's a professor at University of California, San Diego, led a commentary in *Science*, called "Airborne Transmission of SARS-CoV-2." We summarized the findings from the workshop in 300 words, which was published in October of 2020, and we let the CDC know that the publication was coming. In response, they temporarily changed their website information about transmission, but then it was changed back. This was weird.

After that, the next major event was that in the following spring of 2021, there were three papers by various groups about airborne transmission, each published within the same week in major medical journals—*JAMA*, the *Lancet*, and the *British Medical Journal*. At that point, the message really broke through, because these papers were published in highly respected medical journals. The medical community could no longer just ignore it. One doesn't tend to pay attention to journals outside of one's field. Soon after that, the CDC adjusted their website and put airborne transmission at the top of their guide on how the coronavirus is transmitted. Later that year, the WHO finally changed its scientific information as well. But they didn't really publicize that change very well. It's almost like they tried to do it under the radar, on December 23, the day before Christmas Eve.

Where did that resistance at the CDC or WHO come from?

I don't know for sure. My impression is that there was a lot of concern about acknowledging airborne transmission

Timeline of policy on transmission of SARS-CoV-2

2020

January 20 — First known case of COVID-19 in the United States is confirmed, using a sample collected 2 days before in Seattle.

January 23 — Wuhan, China, is placed under quarantine.

January 31 — The WHO declares a public health emergency.

February 23 — The WHO website states that the virus can be transmitted if people breathe droplets from an infected person who coughs or exhales them, recommending distancing of 1 meter (3 feet).

March 11 — The WHO declares COVID-19 a pandemic, tallying more than 118,000 cases across 114 countries, including 4,291 deaths.

March 13 — President Donald Trump declares the pandemic a national emergency.

March 28 — The WHO insists in a tweet that COVID-19 is not airborne, saying, "Droplets are too heavy to hang in the air. They quickly fall to the ground."

April 3 — Lidia Morawska organizes a meeting with the WHO to raise concerns about airborne transmission of COVID-19.

The CDC recommends all people wear masks in public.

July 9 — For the first time, the WHO acknowledges the possibility of aerosol transmission after more than 200 experts, including Marr, write an open letter to them that SARS-CoV-2 is airborne. The WHO continues to emphasize that transmission is through close contact or surface contamination.

September 21 — The CDC removes guidance from its website, posted 3 days earlier, that says transmission of SARS-CoV-2 is airborne and can spread over distances greater than 6 feet.

October 5 — The CDC acknowledges the potential for airborne spread in its guidelines on SARS-CoV-2 transmission, including when 6 feet of distancing has been maintained but the area is poorly ventilated or people are breathing heavily.

The WHO states, "Aerosol transmission can occur in specific settings," at least outside of medical facilities.

March 7 — Johns Hopkins Center for Health Security convenes a National Advisory Committee to develop the Model State Indoor Air Quality Act, which will serve as a template for new state laws.

2023

December 8 — Departments and agencies commit to cleaner indoor air across the nation. The Biden Administration commits to specific actions to improve indoor air quality in federal buildings.

November 18 — The Lancet COVID-19 Commission Task Force publishes ventilation guidelines.

October 13 — The American Society of Heating, Refrigerating and Air-Conditioning Engineers, which sets standards and design guidance for buildings, releases a position statement on infectious aerosols.

October 12 — The White House holds a summit on indoor air quality.

2022

December 23 — The WHO acknowledges airborne transmission.

May — The CDC changes the information on COVID-19 transmission on its website to emphasize that "people can breathe in" droplets and very small particles containing virus that have been breathed out by another person.

April 30 — The WHO for the first time says "aerosols remain suspended in the air or travel farther than 1 metre (long-range)."

April 14–16 — Three major medical journals—*JAMA*, the *British Medical Journal*, and the *Lancet*—all publish articles on the airborne transmission of SARS-CoV-2 in the same week.

2021

December 14 — The first vaccines for COVID-19 were administered in the United States.

because of the implications it would have for health care. In hospitals, if you say a disease is airborne it means you need respirators for all health care workers and airborne infection isolation rooms, meaning negative pressure rooms, for patients. That's very expensive. A lot of hospitals just don't have the capacity to do that for the number of patients that they were getting. And so the WHO did not want, I think, to put hospitals in a bad position, where they would not be able to provide the care that is expected if you have an "airborne" disease.

I now know that the CDC will not make a recommendation for materials if they are not available. Early on in the pandemic the CDC was not going to recommend respirators for people even if they wanted to, because N-95s were simply not available. We didn't even have enough for health care workers at that point; they were sterilizing and reusing them.

I imagine, just seeing how the world works, that liability would be a big issue too, for health care workers who got sick on the job. If the airborne transmission were acknowledged, then the employer is responsible for providing respiratory protection for their employees.

How long did it take to change policy around airborne transmission?
In 2021 the CDC reorganized its guidance for the public about transmission and started talking about ventilation. The WHO had a really nice document on ventilation as well. It was during 2021 that they finally started providing guidance about ventilation and placing less emphasis on wiping down surfaces. But there was still a lot of emphasis on handwashing. That's still the message that a lot of the public remembers: six feet and handwashing.

Then there were, of course, all kinds of political issues about masking. The guidance gradually changed, but people had already heard the initial messages, and I think there wasn't a big campaign to help people recognize the change in guidance. By the time the guidance changed, masks had been politically poisoned. Eventually in 2021 the White House Office of Science and Technology Policy tried to promote awareness of ventilation and virus transmission.

How did you feel going through this?
I was exhausted, because I was working a lot of hours for two years. I was doing 10 media interviews a week for a couple of years, on top of everything else I was trying to do. It was very frustrating to have to fight so hard to try to educate people. The public was open to it, but the medical community—they had learned something in a textbook. It's hard to change your beliefs when you've learned something one way.

What have you learned from this process of policy change?
I've learned that it is messy. It's slow. It's hard to get it right the first time. There are so many different factors and stakeholders. It's not just about following the science, but you have to consider feasibility, the public's willingness to hear a message. If you say something that's not achievable, people aren't going to bother. There are economic considerations. There are political considerations. I realized at one point that maybe a mask mandate is actually counterproductive, because it makes people hate masks so much, or the enforcement is so onerous. There were fights over this, shootings over this. At that point I realized, well, ideally everyone would wear masks, but in our current social and political setting maybe mandating that is not the best thing.

An environmental engineer or aerosol scientist has the right expertise to understand what is happening when a pathogen is moving around among people.

I realized that you have to get the message to decision-makers or key players in the venues that they use and respect. You have to find an ally at some point who's in that community. In terms of getting into the medical journals, it was important to find an ally who was an MD.

What are some common misconceptions you encounter about airborne transmission of COVID, even today?
A lot of people still think that handwashing and distancing are the most important things they can do. But really they need to be thinking about the air we're breathing. That's harder, because we don't see it. We're not used to thinking about that. The message for that came out much later, when people were fatigued. It wasn't promoted as much.

For institutions who want to make sure they're following good guidelines for preventing the spread of COVID-19 now, where can they go for examples?
The U.S. Environmental Protection Agency has good information about buildings and ventilation. It's called the Clean Air and Buildings Challenge. The CDC has some good information now also about masks and the importance of good ventilation. The WHO has a good worksheet on ventilation. The Harvard Healthy Schools and Healthy Buildings program also has some excellent guidance documents available.

What are you working on now?
My lab works on understanding how viruses survive in the environment, in droplets or aerosols. (*See "Optimal Conditions for Viral Transmission," April 3, 2020.*) We've wrapped up our last paper on SARS-CoV-2. We aerosolized virus, pulled it through different kinds of masks, and looked at how much would transfer to your fingers if you were to touch it. We used artificial skin. We didn't find any infectious virus that transferred to skin. So this idea of touching a contaminated mask maybe isn't such a big concern.

Then we have a big interdisciplinary project called Mitigate Flu, looking at transmission in day care centers in Michigan and then also using some animal models to study transmission under controlled conditions. We're trying to develop better methods of detecting a virus in the air and on surfaces.

Then we have another project, doing what's called a human challenge study, where we will deliberately infect people with the flu virus, and then put them in certain situations with other people who aren't infected to see how many people get infected.

And finally, I have a student who's looking at antibiotic resistance, how bacteria and their genes are transported through the air. ■

In this roundup, managing editor Stacey Lutkoski summarizes notable recent developments in scientific research, selected from reports compiled in the free electronic newsletter *Sigma Xi SmartBrief*: www.smartbrief.com/sigmaxi/index.jsp

Universe's Background Noise

A 15-year study found strong evidence of ubiquitous, low-frequency gravitational waves rippling across space. The waves are evidence of large, powerful events, such as interactions between supermassive binary black holes. Researchers from the North American Nanohertz Observatory for Gravitational Waves (NANOGrav) combined observations from three telescopes—the Arecibo Observatory in Puerto Rico, the Green Bank Telescope in West Virginia, and the Very Large Array in New Mexico—to create a galaxy-sized antenna. The huge telescope gathered information from 68 *pulsars*, the ultra-dense remnants of supernovas, which sweep beams of radio waves through space at extremely stable rates. Changes in the timing of the radio wave pulses are evidence that passing gravitational waves are modulating the pulsars' normal rhythms. Albert Einstein predicted the existence of gravitational waves more than 100 years ago, and researchers working on the Laser Interferometer Gravitational-Wave Observatory announced the first evidence of high-frequency waves in 2015. The NANOGrav discovery is significant because low-frequency waves provide evidence of much larger, slower events. Further studies may help resolve individual gravitational events and start to provide new information about the large-scale evolution of the universe.

Agazie, G., et al. The NANOGrav 15 yr data set: Evidence for a gravitational-wave background. The Astrophysical Journal Letters *951:L8 (June 29).*

Sweet Partnership

Honey badgers in Tanzania might cooperate with honeyguide birds in order to acquire their titular food. Decades of unconfirmed anecdotes describe the birds (*Indicator indicator*) leading the badgers (*Mellivora capensis*) to beehives. The badgers would break open the hives to feast on honey and larva, and the honeyguides would enjoy the leftover spoils. Researchers have never witnessed the bird–badger partnership, which has led many to dismiss the accounts as romanticized fiction based on similar cooperative behaviors between honeyguides and people. A team led by ornithologist J. E. M. van der Wal of the University of Cape Town in South Africa tackled the question of the animals' relationship from several angles: literature review, their own observations, and interviews with 394 honey-hunters from 11 communities across nine African countries. The majority of honey-hunters



Wikimedia Commons

had never witnessed the cooperative behavior; however, members of the Hadzabe, Maasai, and mixed cultural communities in Tanzania reported having observed honey badgers and honeyguides interacting. These findings suggest that the behavior might be highly localized, difficult to observe, or both. Future research could include outfitting honey badgers with audio-recording GPS collars to see whether they respond to the birds' calls, or remotely triggering model honeyguides to see how the badgers react. The completed study demonstrates the value of integrating Indigenous knowledge into scientific research.

van der Wal, J. E. M., et al. Do honey badgers and greater honeyguide birds cooperate to access bees' nests? Ecological evidence and honey-hunter accounts. Journal of Zoology *doi:10.1111/jzo.13093 (June 29).*

Cognitive Toll of Long COVID

Patients with lingering COVID-19 symptoms performed worse on memory, reasoning, and motor control tests than control participants. An interdisciplinary team led by epidemiologist Nathan J. Cheetham of King's College London performed cognitive tests in summer 2021 on 3,335 volunteers recruited through the U.K.-based COVID Symptom Study smartphone app; of those volunteers, 1,768 also participated in a second round of testing in spring 2022. Participants with active COVID-19 infections performed worse on the first round of testing than healthy individuals, and those whose symptoms lasted for 12 or more weeks had notably lower scores than those who had recovered more quickly. In the second round of testing, individuals who had short or mild cases of COVID-19 no longer displayed cognitive deficits, but those with long, severe, or lingering symptoms did not improve in the testing. The group's results are preliminary and are hampered by a cohort that was strongly skewed white, female, and middle-aged. Nonetheless, the study adds to the evidence that some individuals may suffer long-term cognitive impairment well after they stop showing signs of active COVID-19 infection.

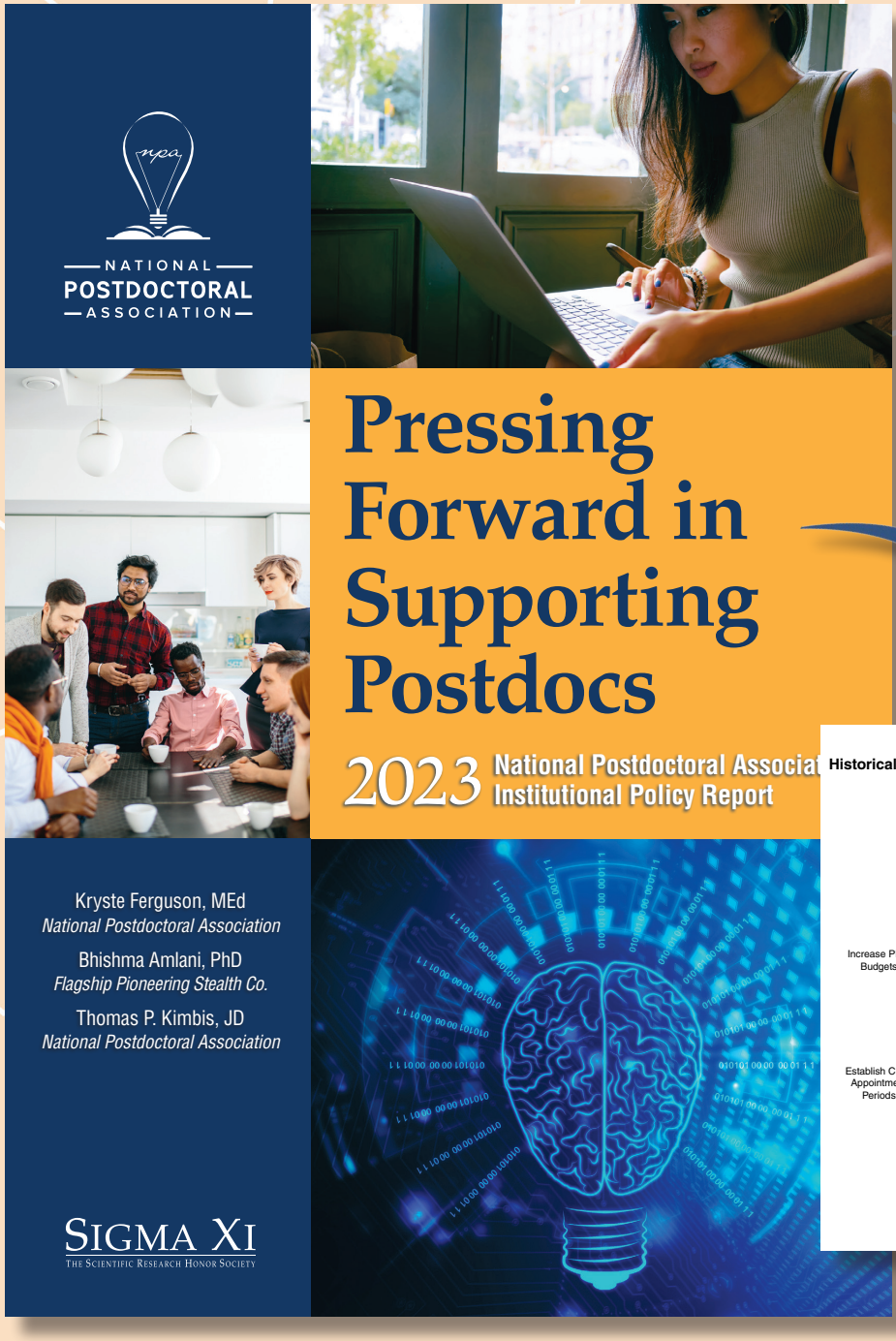
Cheetham, N. J., et al. The effects of COVID-19 on cognitive performance in a community-based cohort: a COVID Symptom Study Biobank prospective cohort study. eClinicalMedicine *102086 (July 21).*

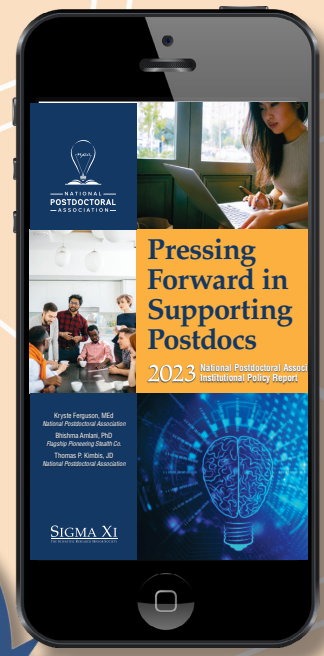
Prehistoric Female Hunters

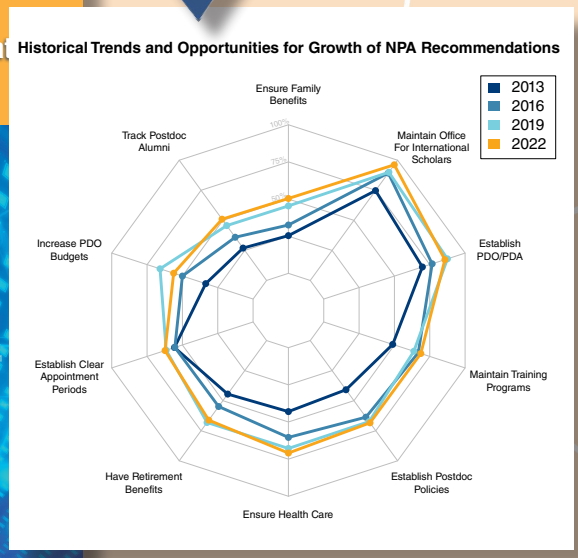
Archaeological and ethnographic evidence has upset the traditional narrative of a sexual division of labor in forager societies. A team of biologists at Seattle Pacific University examined reports of hunting practices worldwide in modern hunter-gatherer societies and in burial evidence in ancient societies. They found that out of 63 foraging societies with clear hunting practices, 79 percent included female hunters. In some groups, such as the Aka in the Democratic Republic of the Congo, ethnographic reports describe women hunting while carrying infants, which contradicts the idea that childcare limited women's movements and forced them to restrict themselves to foraging. Assumptions about women's activities have also distorted the ways that researchers have interpreted archaeological sites. For example, a Viking burial ground excavated in the 19th century included an individual buried alongside weapons associated with high-ranking warriors. Researchers assumed that the individual was male until a 2017 genomic study confirmed that she was female. The Seattle Pacific University team's analysis of the data from historic and modern hunter-gatherer groups on all populated continents suggests a broader need for researchers to rethink their ideas of traditional societies.

Anderson, A., S. Chilczuk, K. Nelson, R. Ruther, and C. Wall-Scheffler. The myth of man the hunter: Women's contribution to the hunt across ethnographic contexts. PLoS One *18:e0287101 (June 28).*

Comprehensive Report on Postdoctoral Landscape Now Available





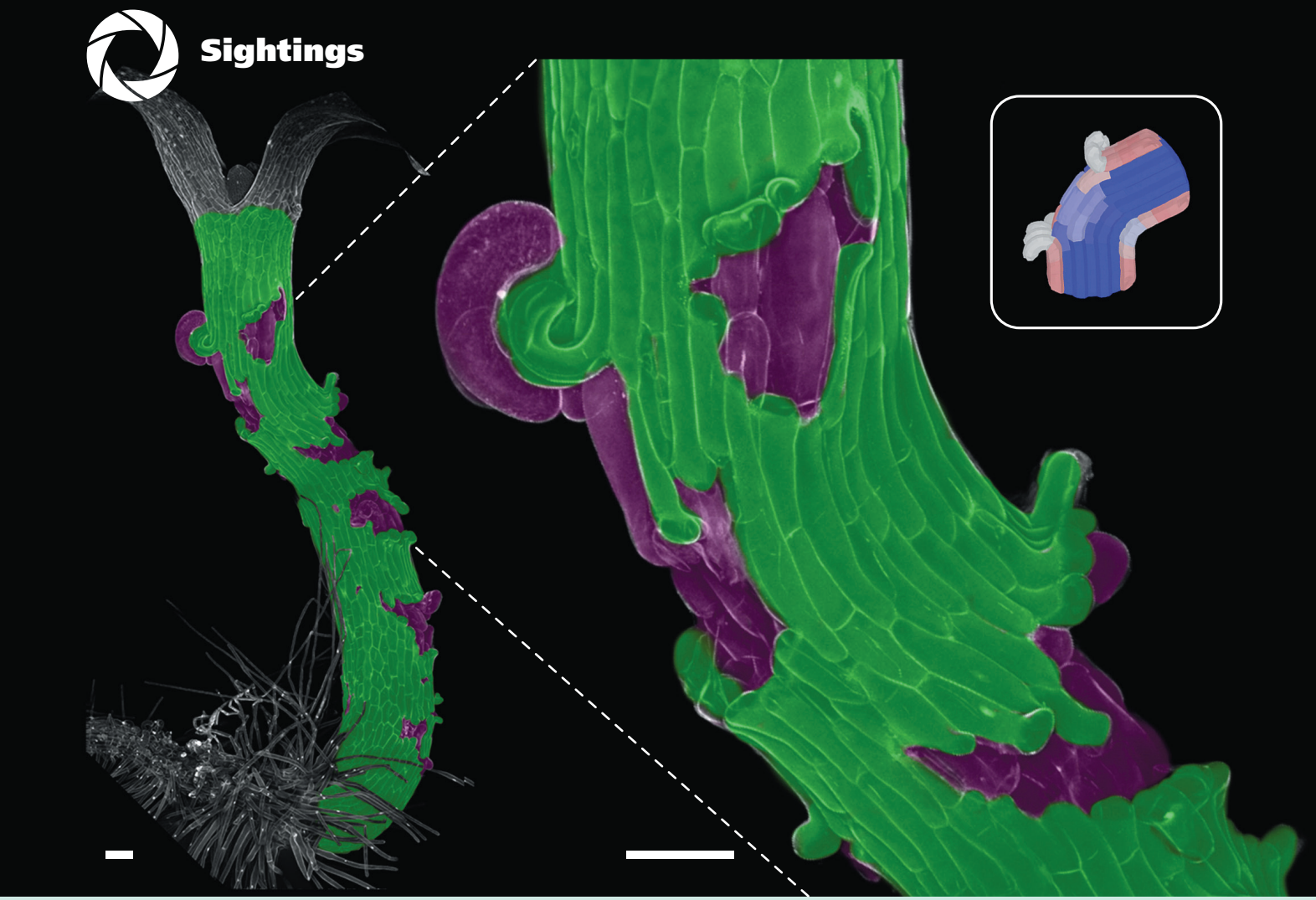


A digital version of the 2023 NPA Institutional Policy Report is available at https://www.nationalpostdoc.org/policy_report_databa or sigmaxi.org/NPA23-Report

The report documents postdocs' institutional training environments and assesses the supporting policies and training programs led by the offices of postdoctoral affairs that direct and manage their individual postdoc communities.

NATIONAL POSTDOCTORAL ASSOCIATION

SIGMA XI
THE SCIENTIFIC RESEARCH HONOR SOCIETY



Images courtesy of Karen Lee and Robert Kelly-Bellow; inset simulation by Richard Smith

Pushing and Pulling in Plants

Stems use mechanical forces to control growth between their layers of tissues.

Relatively few organisms are a single cell thick, so their layers of cells have to work together to build a functional structure. Genes in different tissue layers are known to influence one another's activity, and most of the time, that communication is assumed to occur via chemical signaling. But because the layers are in contact, there's another option: mechanical forces. "We think of a biochemical signal as the analogy of verbal communication: You want to get somebody to do something, you tell them to do it. But tissue layers interacting with each other via pure mechanics is like me grabbing somebody and moving them to where I want to do something," says Robert Kelly-Bellow, a plant biologist at the John Innes Centre in the United Kingdom.

Kelly-Bellow, working with plant scientist Karen Lee and their colleagues in Enrico Coen's laboratory, stumbled across a model organism to test the role of mechanics in plant stems as they grow. The plant—a tiny carnivorous aquatic species called *Utricularia gibba*, a type of bladderwort—has a stem made of a central vasculature surrounded by an array of blades with air pockets between them, then an outer wall or epidermis. The research group was trying to grow a mutant to study the plant's modified leaf cups, which trap microscopic prey. But they ended up with a super-deformed dwarf with a thick epidermis and a twisted stem vasculature. What was making the central structure warp that way?

The group then added computational modeling, led by computer scientist Richard Kennaway, to their study, and showed that if the growth rate of the epidermis was lowered—a slower growth rate being quite possible with increased thickness—then the inner layers with their normal growth rate would be compressed and buckled as they grew. The mechanics alone would account for the structural difference (*see figure on page 273*). "If the outer layer is being affected more strongly, it's acting like a sort of straitjacket

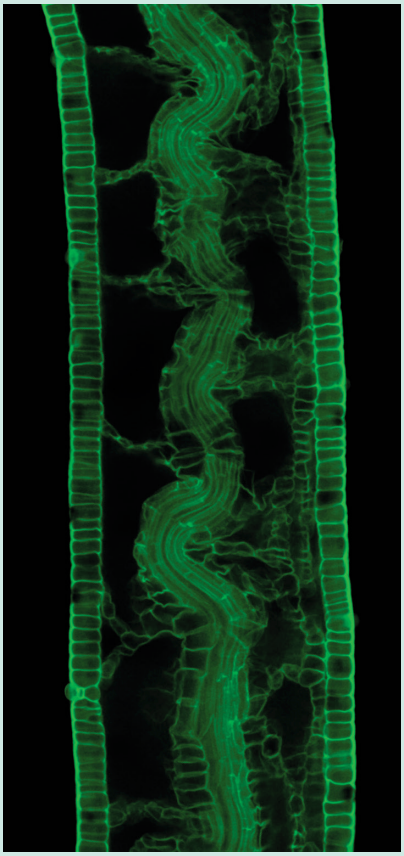
One way to learn how a process works is to try to break it. That is what's happening in this thale cress plant (*left*), which researchers bred as a dwarf variety with a thick epidermis (*green*) and weaker adhesion between cells. Internal tension and compression in the plant increase during growth; small cracks become large tears to relieve these forces (*inset simulation*). This mutation shows the role of mechanical forces between tissue layers in the plant.

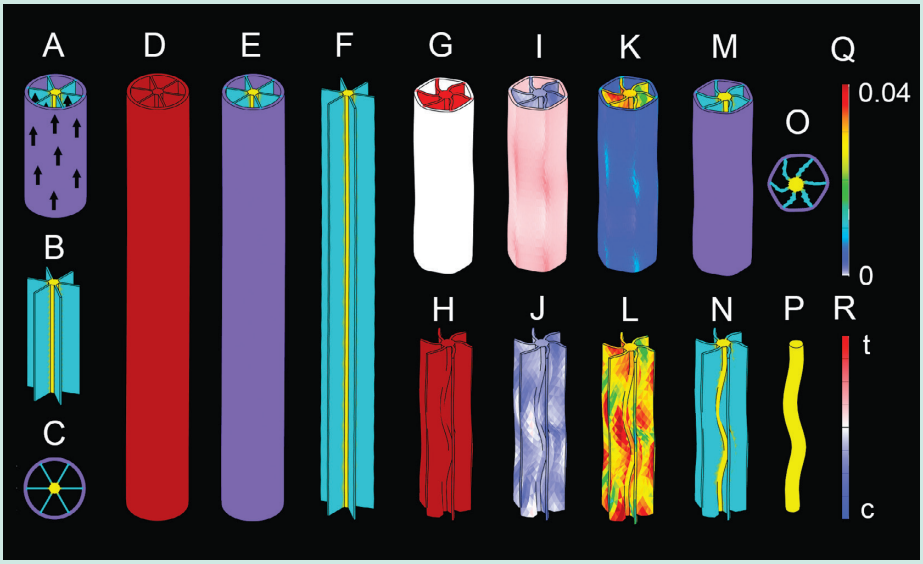
that's tightening during development," Kelly-Bellow says. "This conflict has to be resolved in some way, and the resolution here is bending and twisting around into the empty space that it has."

The group then explored the role of a plant growth hormone called *brassinosteroid*. Kelly-Bellow and his colleagues mapped a gene to show that this hormone is not properly synthesized in the dwarf plant. But if the dwarf plant is treated with this hormone, it begins to grow normally. The hormone seems to have a role in reducing the stress between tissue layers, in essence loosening the epidermis, although the exact mechanism is still unknown. It could be reorienting the growth of structures such as microfibrils or microtubules in plant cell walls, which can in effect change the tissue stretchiness. "Maybe without brassinosteroid, it's producing more random arrays of microtubules, as opposed to ones with certain alignments, adding to wall thickness," says Kelly-Bellow. "An alternative hypothesis is that it might have to do with microfibril interactions, the sliding between layers. Maybe adding brassinosteroid preferentially loosens the fibers in some way, allowing the shape and alignment of the cell to expand more quickly."

To extend their hypotheses to other plants, Kelly-Bellow and his colleagues then moved onto *Arabidopsis thaliana*, also known as thale cress, a small weed in the mustard family that is often used as a model organism because of its relatively small genome (it was the first plant to have its genome fully sequenced). A mutated version of this plant has weakened adhesion between cells, and Kelly-Bellow and his colleagues showed that if it is grown with a compound that inhibits the plant's ability to synthesize brassinosteroid, it develops more pronounced stem cracks. The group then crossed this mutant with a dwarf line that was brassinosteroid-deficient. They found that this double-mutant grows with very dramatic epidermis splits and cracks (*see figure on opposite page*). In another model at the cellular level, led by computational biologist Richard Smith, the team showed that this splitting changes the internal stress in the plant stem (*inset on opposite page*). These results showed that altering the mechanical forces between the layers could have great effects on growing plant stems. "The mutant dwarf plant has this increased role of the straitjacket increasing tension on the outer layer and producing increased compression on the inner layer. Now, if you start to cut holes in that straitjacket, they just rip apart because this conflict between inner and outer layers is relieved," Kelly-Bellow says.

Understanding the role that mechanical forces play in tissue layer growth, and how hormones can regulate those interactions, can help unravel the general process of organism growth. There are implications for agriculture, such as in efforts to develop more drought-tolerant but healthy dwarf varieties of crops. Similar mechanical forces also are known to be significant in some animal tissues, such as in the cracking texture of crocodile skin or the shaping of the fingerlike projections called *villi* in the intestine. "Something that's key for me is to show that there is another way that these tissue layers interact with each other instead of just thinking about things in a signaling way," Kelly-Bellow says. "We only think about brassinosteroid as a hormone, so some sort of chemical must go from A to B. Well, maybe there's a signal-less way of doing things." —*Fenella Saunders*





Images from R. Kelly-Bellow et al., 2023, *Science* 380:1275; simulation by Richard Kennaway

The tiny water plant *Utricularia gibba* usually has a straight central vasculature surrounded by blades and an epidermis (*shown in an idealized simulation as yellow, cyan, and purple at A–C and E–F above*). But when the plant is mutated into a dwarf with a thick epidermis, its vasculature becomes twisted (*confocal microscope image at left*). In a simulation, the normal growth rate (*shown in red at D; growth scale at Q*) in the dwarf plant drops to zero just for its epidermis (*white, G*), and the stem's internal structure—growing normally—begins to buckle under the constriction (*H*). This twisting growth puts the epidermis in tension (*I*) but relieves some of the internal compression (*shown at J, stress scale for tension [t] and compression [c] at R*). The resultant growth rate is low for the epidermis (*K*) and variable inside (*L*), resulting in the warped vasculature (*shown at M–P*). The simulation demonstrated that mechanical forces could explain these tissue interactions.

Perspective

Direct Detection of Exoplanets

An optical technique that uses the wave nature of light could reveal planets outside our Solar System more accurately.

Marija Strojnik

The first time astrophysicists detected a planet outside of our Solar System around a Sun-like star was in 1995, and since then, more than 5,000 extrasolar planets (or exoplanets) have been identified. Despite that impressive number, detecting exoplanets is an incredibly difficult task. For an exoplanet the size of Jupiter, the planet is about a million times fainter and a hundred times smaller than the star it orbits. There is currently no optical instrument on Earth or in space that can separate these two objects as individual entities. Instead, the existence of such an exoplanet can be inferred from various data, but no Jupiter-equivalent exoplanet has yet been directly detected. We can assess its distance from its star and its temperature, but very little other information.

So far, exoplanets that are giant or that orbit very close to their stars (or both) have been the most readily detected. To systematically find exoplanets that more resemble Earth in size, temperature, and distance from their star—thus increasing their likelihood of being habitable—will require innovative approaches. Research that I have conducted from theoretical calculations in the 1990s to a proof-of-concept experiment in the past few years considers how to use the wave nature of light as a tool for direct detection of exoplanets, potentially with greater accuracy than current methods permit. Although this technique could take several decades to be proven and implemented, it has the advantage that the signal used can be present only when an exoplanet exists, reducing the possibility that anomalies will affect the detection process.

How to Search for an Exoplanet

Exoplanets cannot be detected visually, so astrophysicists began to look for other measurable quantities that can be detected in the physical characteristics of a star and that create an anomaly that would likely not exist if the star did not have orbiting planets. All the current techniques of exoplanet detection have been adapted from radiometric measurements, in which a temporal dependency is introduced because the planet orbits its star and thus provokes a measurable change in some of the star's parameters.

Arguably, the most productive planet detection technique has been what's called the *transit method*. A dim, small object orbiting a bright, large object will decrease the amount of light that an observer receives from the bright object only while the dim object travels in front of it. This dip in the detected power is then related to the planetary orbit and other parameters of celestial mechanics. The downside of this technique is that it is using a small, indirect signal, which could be confused with some other anomalous signal from the star instead of coming from a planet.

Another interesting technique of planet detection is the *radial velocity technique*. In a planetary system, we can see only the star. For half of the *local year*—the time that the planet travels around the star once—the planet is traveling toward an observer, and the other half of the local year it is moving away from an observer. The relative speeds of the planet are highest and lowest when the planet is at its farthest distance from the star, as seen by the observer. This relative change in speed means that when the planet moves away from the observer, the spectrum of the electromagnetic radiation it emits is shifted toward longer wavelengths. Conversely, when the planet travels toward the observer, its spectrum shifts toward shorter wavelengths. The time between the high and low peaks in the planet's spectrum tells us the length of the local half-year, and the amplitudes of the peaks can be used to deduce the planet's temperature. However, the temporal separation between the two measurements—of half of the local year—makes such radiometric measurements very demanding. In half an Earth year, the precision of the detection instrument could deteriorate by more than the power of the signal from an exoplanet. Thus, this method has mostly been used to find planets that orbit close to their stars and therefore have short local years.

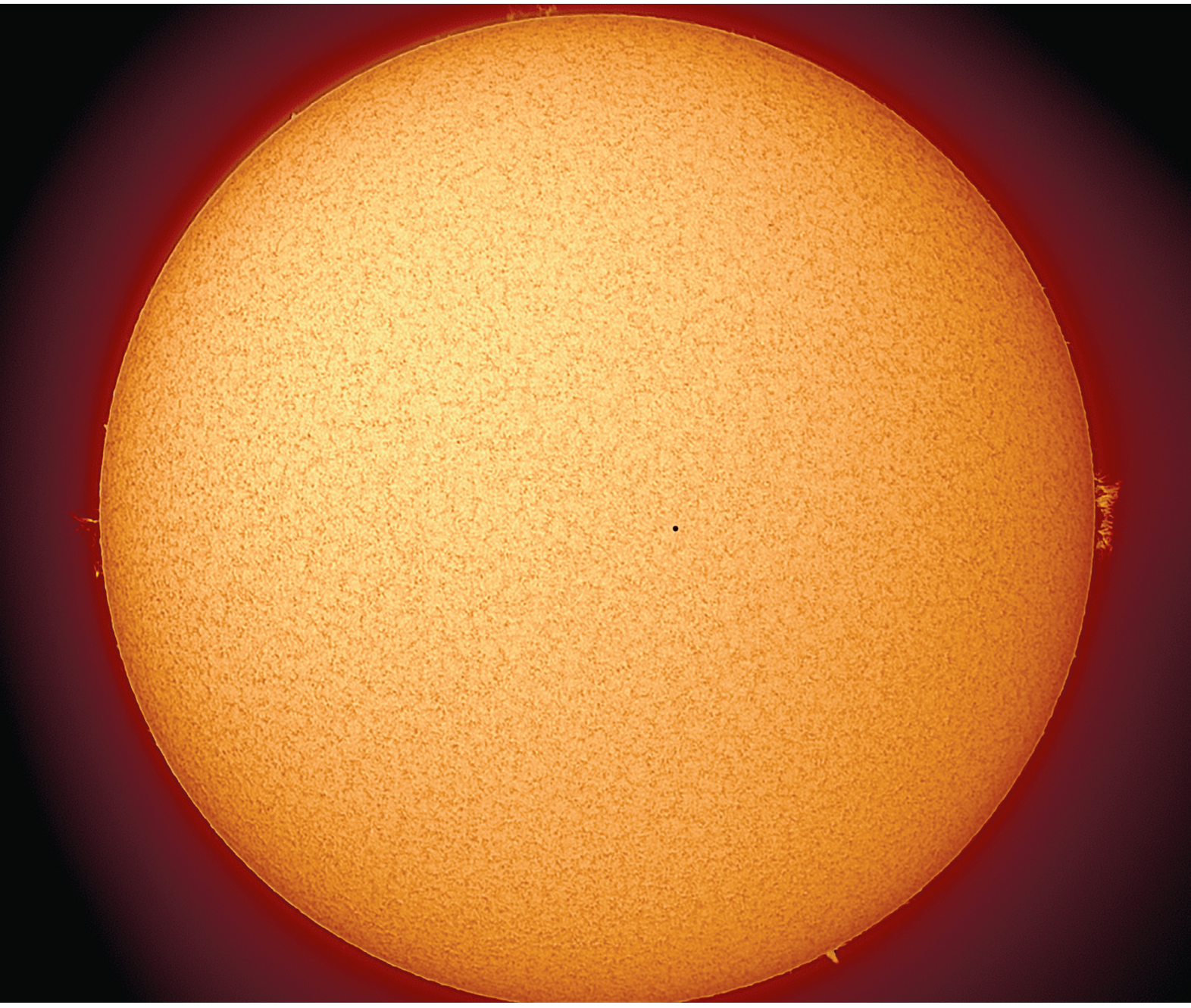
Astrometry is a technique that uses interferometry to measure the distance to a star with unknown coordinates with respect to a known star. *Interferometry* is similar to throwing two rocks into a calm pool of water, creating two sets

QUICK TAKE

Thousands of exoplanets have been identified using several different methods, which mostly use the amount of light, radial velocity, or "wobble" of those planetary systems.

All of these methods currently have drawbacks, including a large time separation between measurements, which introduces the potential for errors.

A method adapted from optical testing uses the wave nature of light to allow measurements to be made in immediate succession, potentially increasing accuracy.

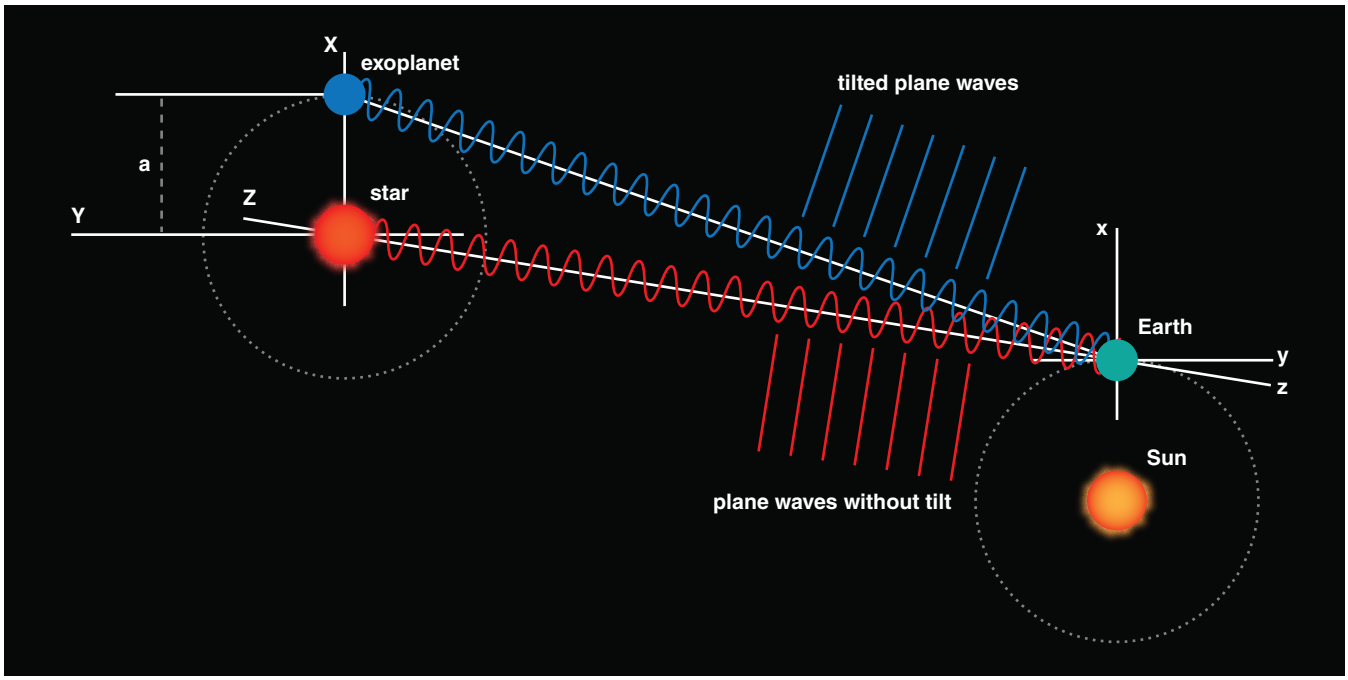


John Chumack/Science Source

In 2019, Mercury *transited* the Sun, meaning it crossed in front of the star as viewed from Earth. In this image, the tiny planet is a dot that can be seen just below right of center of the Sun. Transits are used to identify extrasolar planets around their stars, but the measurements are indirect. An optical method using the wave nature of light could provide a means of direct detection.

of concentric circular waves until the waves meet and the pattern breaks. In astrometry, optical radiation is collected from a reference star and the unknown star and then interfered. The resulting pattern is a set of parallel straight lines, called *fringes*, directed perpendicularly to the line connecting the stars. The distance between fringes is inversely proportional to the star separation, meaning the larger the separation, the smaller the distance between two neighboring fringes, and this distance can be used to calculate the separation between these two celestial objects.

When the unknown star has a planet, the visible star behaves as a variable star, rotating around the common center of mass of the star and planet. The detected interference pattern changes in time and goes through a complete cycle during a local year, when the planet completes one orbit around its star. When the invisible planet is between both stars, the center of mass of the planetary system moves closer to the reference star, and the interference pattern presents fringes that are slightly more separated. When the planet is on the same line as both stars, but distant from the reference star, the center of mass of the planetary system moves away from the reference star, and the interference pattern presents fringes that are slightly less separated. This change in fringe separation over time could be considered a *derivative*, a technique that is used to measure a small signal buried under a large, constant level of noise. However, this technique is also hampered by requiring two precision



Adapted from Marija Strojnik

Light waves that travel from other stars and their exoplanets originally start out as spherical waves, but after they travel long distances to reach Earth, these light signals will become plane waves. The light wave fronts originating from the orbiting exoplanet will be inclined in comparison with those from its star, because the planet is located at an off-axis position. The different orientations of the two wave fronts can be used to create an *interference pattern* that shows the existence of the exoplanet.

measurements separated by a large time interval, creating room for error. Nonetheless, the European Space Agency's Gaia mission has recently started detecting exoplanets using this technique.

Finally, there is the imaging technique called *coronagraphy*, discovered about 100 years ago when scientists were trying to understand the Sun and its corona. The bright Sun tends to overwhelm its corona, so this technique uses a black mask to occult the Sun while its corona alone is observed. Similarly, in the search for exoplanets, an occulting aperture is used to block the star's radiation while its planet is observed. This technique has been studied for exoplanet detection only in the past few years, and it is best for widely separated, young, high-mass planets. There are upcoming programs that plan to utilize this technique.

A New Search Technique

My planet search story began at the Jet Propulsion Laboratory (JPL), managed for NASA by the California Institute of Technology. JPL is tasked with robotic spacecraft exploration of the Solar System. Although most of the instruments in their robotic exploration are controlled from Earth, each robot's optical navigation uses an onboard autonomous computer system that incorporates a library of stars and an algorithm. In 1992, I demonstrated this system's successful performance at an observatory as a technology development project, in support of the Cassini mission to Saturn, which launched in 1997 and ended in 2017.

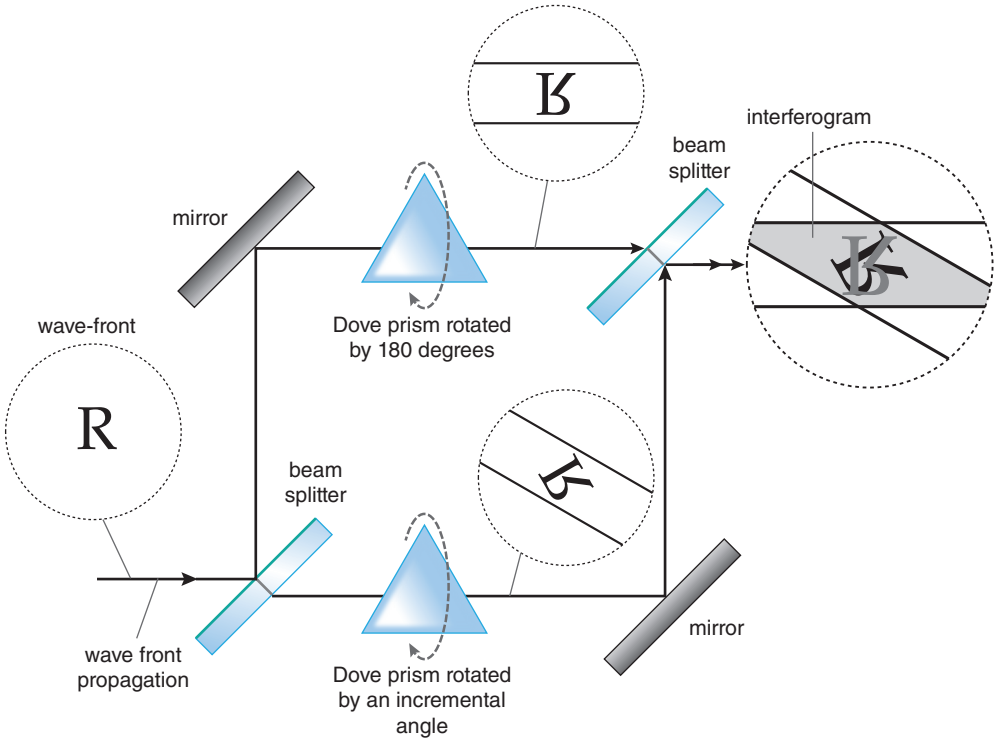
As JPL analyzed the robotic images from the objects in our Solar System, the logical next step was for their scientists to assess the possibility of searching for planets outside our Solar System. That problem was much more difficult than examining a planet next door, and it included several almost-impossible-to-solve technological problems, a truly delicious puzzle for scientists: resolution, distance, dynamic range, signal-to-noise ratio, and scattering, to name just a few. I became involved in this work at JPL as an optical scientist with a specialization in infrared and instrumentation. My expertise gave me a different perspective from planetary scientists. I began working on the theoretical foundations of an optically based technique in the 1990s, which was developed into a proof-of-concept model in 2019.

When analyzing successful and promising planet detection techniques, we learned that several of them take an imperfect derivative of a signal: Long time separations are involved, allowing for the possibility of systematic errors, ghost images, and degradation because of aging.

Planetary transits also illustrate the type of noise, or spurious signal, that may creep into planet detection schemes because no star, except the Sun, can be resolved with any existing instrument on or close to Earth. For any other star, the decrease in its detected power because of a dark pixel that represents a relatively cold planet could be simply canceled out by noise from prominences or other types of powerful star emissions.

The other useful lesson from astrometry was that we really do not need to obtain the whole image of the searched object, because two objects can create a single signal. The presence of the second object can be established simply by detecting an interferometric signal. It occurred to me that a technique used in testing optical surfaces could be adapted to sensing two objects by creating one signal, but without the time delay between signals that allowed for the introduction of errors.

Outside of its use in astrometry, interferometry is also a common technique used to test the quality of optical surfaces in telescopes: By interfering light beams that travel to the test surface and to a high-quality reference surface, optical engineers obtain intensity distributions that they relate to the quality of the surface being tested. But within the



Adapted from Marija Strojnik

A key element in a rotational shearing interferometer is a *Dove prism*, an optical device that can be rotated to change the orientation of the wave front that passes through it. A test wave front, here the letter R, is split into two beams that pass through different Dove prisms. The top Dove prism rotates the R by 180 degrees. The bottom Dove prism can be rotated by incremental angles to change the wave front. When the two beams are recombined, they create an interference pattern, or interferogram, that can be used for optical detection.

past 50 years, we have seen a rapid development of the optics industry that has been so successful that we have run out of suitable, easily fabricated reference surfaces to use in tests. The solution has been to use *self-referencing* for testing. If the test surface is measured, then displaced slightly along one axis, then remeasured, the two measurements produce a derivative. Using mathematical integration, we can recover the surface shape and determine deviations from the shape's required value. This operation of displacing a surface with respect to itself is called *shearing*.

Besides displacing a surface along a single direction, one can also rotate a surface. Most traditional optical systems, and therefore their constitutive components, exhibit rotational symmetry around the optical axis, which in this case would be the axis of symmetry.

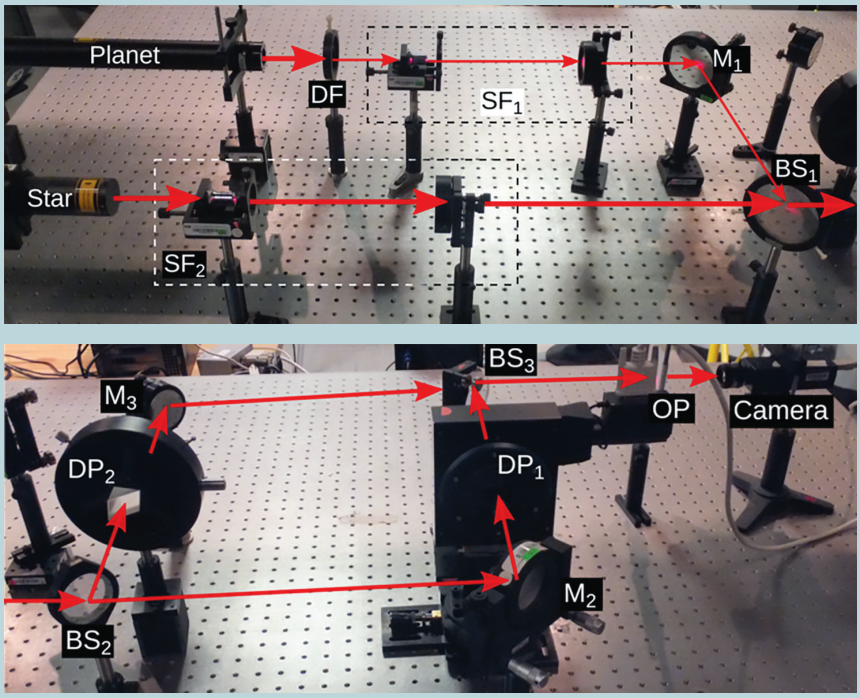
If we wanted to test whether a surface is actually rotationally symmetric, we could optically couple it to a *rotationally shearing interferometer* (RSI). We start out with what's called a *Mach–Zehnder interferometric configuration*, in which each beam travels along each path only once. We transform it into an RSI by inserting what's called a *Dove prism* into each interferometer arm; one of the Dove prisms is rotated so that it also rotates the wave front, whereas the other wave front passes through unaffected (*see figure above*). When the beams are combined, the wave fronts are superimposed and subtracted, generating an interference pattern that contains the information about how similar the rotated surface is to the original surface. We increment the orientation angle of the Dove prism by a few degrees in a series of measurements until the rotation angle of 360 degrees is achieved. At that time, the surface is compared again to the original surface.

If there is a planet, it will introduce asymmetry into the optical measurements.

Let's consider a surface that is perfectly rotationally symmetric, except for a bump a few degrees to the right just below center—as was the case for an image of the Sun on November 11, 2019, when Mercury *transited* the Sun, meaning it traveled in front of the Sun as viewed from Earth (*as shown on page 275*). If this surface is rotationally sheared and measured, all the subtractions between successive surface positions for different orientation angles would result in a signal at the bump location. The bump would trace a radial anomaly. Therefore, my group at the Optical Research Center in León, Mexico, had the idea of using an RSI to look for anomalies with stars that indicate exoplanets, rather than surface aberrations in optical surfaces.

A lone star is a perfect example of a rotationally symmetric optical system. Outside the Earth's atmosphere, a star looks like a flat white circle, because of the isotropic nature of its radiative emissions. If you rotate an image of the star alone by a few degrees and subtract its emission characteristics from those of the original, you would end up with zero.

The situation is appreciably different when we are dealing with detecting a distant star and its planet with an RSI located on Earth or in its orbit. The instrument will focus on the star because its overwhelming brightness is all we can see. But if there is a planet, perhaps in orbit at the distance of Jupiter, it will introduce asymmetry into the measurements. We can use the wave nature of light to detect the planet. The spherical waves that originate at the star and its planet separately, after traveling the long distance to the Earth, become plane waves. But the wave fronts originating from the planet will be inclined, because the planet is located at an off-axis position (*see figure on page 276*). The star's wave front also includes some inclination angles, but they are all nearly



A laser setup tests the concept of detecting a planet around a star with rotational shearing interferometry. The beam from the "star" is on axis with the system, whereas a beam from a "planet" is offset (by mirror M_1) as it would be in orbit. Laser beam paths are shown by the red arrows. Laser beams first go through a set of filters (shown by DF, SF_1, and SF_2) and beam splitters (BS_1 and BS_2), so the combined light of the star and the planet goes through two Dove prisms (via mirror M_2 to DP_1 and to DP_2). The Dove prism at DP_1 is rotated, introducing a wave front shear in the signal. After the beams are recombined (via mirror M_3 to BS_3) and the output (OP) is imaged by the camera, the derivative shows interference fringes that indicate the planet. (Images courtesy of the author.)

zero. A filter that eliminates small-angle plane wave spectra will leave only the planet information.

To implement this filtering with rotational shearing interferometry, we subtract one complete wave front that is incident on the interferometer from the same wave front that has been mechanically rotated using a Dove prism. This action results in a destructive interference for all positions on the aperture except where the planet is located. There, we detect an inclined wave front, which in the RSI manifests itself as straight fringes, just as in astrometry. But the fringes are recorded during a single measurement, taking less than a few minutes.

Our technique of exoplanet detection therefore addresses the challenges of eliminating the elapsed time between consecutive measurements, avoiding systematic errors, and eliminating the validation of results using statistical means. It takes a derivative with respect to the angle corresponding to the orientation of the planet with respect to its star, so when there is no planet, there will be no fringes.

Putting the Technique to Work

We know that two-aperture interferometry works well, because it has been successfully implemented in astrometry to measure star positions. With the RSI, we have a faint planet instead of the second star. Fringes are visible only if the planet is within the field of view of the instrument. But both our simulation studies as well as our theoretical analysis indicate that the density of fringes (the number of fringes per unit distance) and their orientation change with the change in the Dove prism orientation. Thus, the planet's presence may be further confirmed by changing the Dove prism orientation angle, and the existence of a fringe pattern may be directly and causally related to the presence of a planet.

The existence of a fringe pattern may be directly related to the presence of a planet.

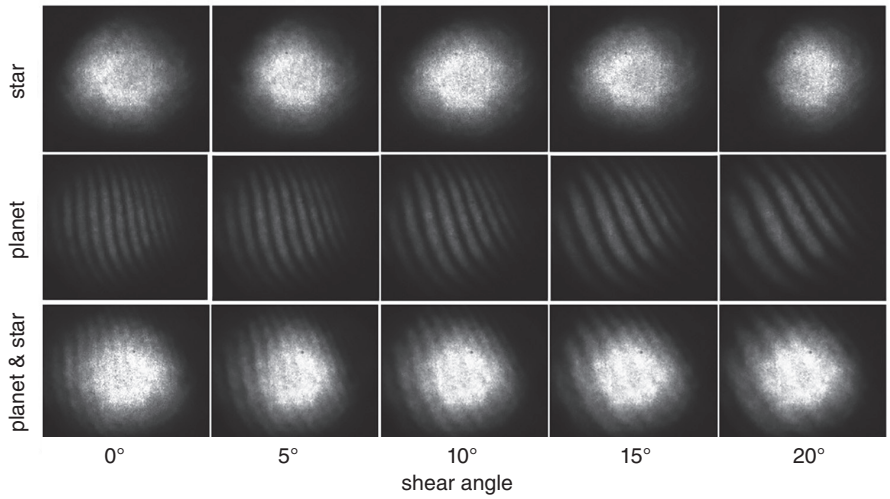
In 2019, we built and tested the first concept demonstration for the RSI. A simulated planetary system made from two lasers is coupled to an RSI (*see the figure at left*). The star laser beam and its beam-conditioning components are aligned to the RSI axis. The planet laser beam is placed at a slight angle with respect to the star laser beam. The planet laser beam is coupled to the star laser beam with a cube beam combiner.

A cube beam splitter divides the combined incident wave front into two beams and sends them into the two interferometer arms. A Dove prism in each arm transforms the interferometer into an RSI. A rotated Dove prism rotates the wave front in one interferometer arm (the horizontal one after the beam splitter) with respect to the beam in the other arm (the vertical one after the beam splitter). An identical, stationary Dove prism in the reference arm compensates for the changes in the optical path. The beams traveling through two arms of the RSI overlap once again at the second beam combiner, and the interferograms are captured by a camera.

The interferograms obtained with the experimental setup when the angle of orientation of the Dove prism increases from 0 to 20 degrees are displayed in the figure on the opposite page. We considered three cases: only the star laser is on, only the planet laser is on, and the complete planetary system simulator is on.

In the first row, the interference patterns remain unchanged when the angle of orientation of the Dove prism increases from 0 to 20 degrees for a single point source (star laser) on the optical axis. This series of interferograms confirms that the RSI is insensitive to a solo star located on the optical axis. A bright field is detected for all angles of the orientation of the Dove prism. No fringe pattern rotation is observed when the incident star wave front possesses rotational symmetry. This is the case for a star without a planetary companion when the instrument is aligned with the star center.

In the second row, the interference patterns change when the angle of orientation of the Dove prism increases from 0 to 20 degrees for a single point source (planet laser) placed at an angle with respect to the optical axis. The practically straight interference fringes decrease in density and increase in inclination angle when the shear angle is increased. This is one of the characteristics of the RSI



Results of testing a proof-of-concept system for rotational shearing interferometry show the different outcomes if the simulation involves a light signal from only a star, only a planet, or a planet with a star. The interference fringes only appear if a planet is present, and confirm that the fringe density and inclination angle change with the shear angle. (Image courtesy of the author.)

when measuring the tilted wave fronts. These experimental data confirm that an off-axis source may be detected in an RSI by generating straight fringes. Furthermore, the fact that the fringes arise from an off-axis source, rather than because of an artifact, may be confirmed by changing the shear angle.

Likewise, in the third row, the interference patterns change when the angle of orientation of the Dove prism increases from 0 to 20 degrees for a complete planetary system (a bright point source on-axis and a weaker point source off-axis). These interferograms are similar to the planet-only interferograms, except that the star incidence provides the background in the central portion of the interferograms. This is a consequence of the Gaussian shape of the laser incidence. Filtering the constant background or mirroring bright and dark pixels would eliminate these effects.

Our experimental data confirm that a dim off-axis source next to a bright source on-axis may be detected in an RSI. The dim source generates faint fringes on top of the bright star incidence, as predicted by theory. The theoretically predicted change of the fringe inclination angle with the shear angle is also confirmed, making the RSI as a planet detection technique fault-tolerant to a spurious signal. The theoretically predicted change in the fringe density is also confirmed in the experiment.

The next step in our proof-of-concept experiment is to build engineering models, have a demonstration in an observatory environment, and further increase design optimization. Then, NASA or some other space agency may decide to adopt the technique. At that time, it will take at least a decade to incorporate the RSI planet detection instrument into a mission. Until we reach those later demonstration stages, we won't be able to prove for certain that an RSI can discriminate better than current instruments between planets and the other miscellaneous information noise. But so far, the results have been promising. I believe that the time has come for us to start building instruments that can speed the measurements of exoplanets using signals that intrinsically are present (or absent) if the planet is there (or not). No other planet detection technique has such strong fault-tolerance and causality built into its detection scheme.

Bibliography

Bravo-Medina, B., M. Strojnik, A. Mora-Nuñez, and H. Santiago-Hernández. 2021. Rotational-shearing-interferometer response for a star-planet system without star cancellation. *Applied Sciences* 11:3322.

Gonzalez-Romero, R., M. Strojnik, and G. Garcia-Torales. 2021. Theory of a rotationally shearing interferometer. *Journal of the Optical Society of America A* 38:264–270.

Scholl, M. S. 1996. Infrared signal generated by a planet outside the solar system discriminated by a rotating rotationally-shearing interferometer. *Infrared Physics and Technology* 37:307–312.

Strojnik, M. 2022. Development of a rotational shearing interferometer for detection of (habitable) exoplanets. *Proceedings of SPIE 12170, Advances in 3OM: Opto-Mechatronics, Opto-Mechanics, and Optical Metrology*, 121700C.

Strojnik, M. 2022. Rotational shearing interferometer in detection of the super-Earth exoplanets. *Applied Sciences* 12:2840.

Strojnik, M., B. Bravo-Medina, and R. R. Baltazar-Barron. 2019. Rotationally-shearing interferometer: Preliminary results with a simulator Solar System. *Proceedings of SPIE 11128, Infrared Remote Sensing and Instrumentation XXVII*, 1112816.

Strojnik, M., and G. Paez. 1999. Simulated interferometric patterns generated by a nearby star–planet system and detected by a rotational shearing interferometer. *Journal of the Optical Society of America A* 16:2019–2024.

Strojnik Scholl, M., and G. Paez. 1999. Cancellation of star light generated by a nearby star–planet system upon detection with a rotationally-shearing interferometer. *Infrared Physics and Technology* 40:357–365.

Marija Strojnik is a distinguished professor at the Optical Research Center in León, Mexico, and is the current president of Sigma Xi, The Scientific Research Honor Society. Email: mariasigmaxi@gmail.com



Unearthing a Prayer for the Dead

Terahertz waves uncover the hidden inscription on a 16th-century lead cross.

David S. Citrin, Alexandre Locquet, and Junliang Dong

Back in 2017, a researcher from the Laboratory of the Archaeology of Metals in Jarville-la-Malgrange, France, brought a 16th-century lead cross to our labs, asking if we could read a barely visible inscription underneath a layer of corrosion. We were intrigued but dubious. Our team, based at the Georgia Institute of Technology, was developing imaging technology that was primarily intended for industrial applications. This request was not the first time that an archaeological lab had asked us to examine an artifact to see whether we could reveal what was hidden below the surface, but many of our past attempts had not been very successful.

We already knew that our imaging technique, which uses *terahertz electromagnetic radiation* (similar to microwaves, and sometimes called *submillimeter waves*), did not work well on corroded iron artifacts. The main chemical constituents of rust are electrical conductors, which terahertz waves do not significantly penetrate. (Previously we had tried unsuccessfully to find the silver filigree work on a 1,500-year-old Merovingian buckle, within what looked like a lump of rust.) This cross seemed somewhat more promising because it was composed of lead, whose major corrosion products are electrical insulators, which allow terahertz waves to pass through. We still weren't optimistic, but we did what we could.

As we feared, there was little obvious to see in our preliminary scans, so we put the data on the shelf (actually on a hard drive) for a few years and moved on to other problems. But the story did not end there. Three years later, after building up our data-processing skills, one of us (Dong) recalled the information sitting on the hard drive and had a new idea about how it might be analyzed. His idea proved successful: Our refined terahertz imaging process enabled us to see through the corrosion layer on the 16th-century cross and to read the prayer inscribed below the corrosion layer. The results offer a new way to analyze archaeological artifacts, and, in this particular field of study, to gain insights into medieval attitudes toward death and Christian faith.



Junliang Dong

When this 16th-century lead cross was excavated in France in 1904, researchers noticed faint traces of an inscription, but they could not decipher the words. More than a century later, terahertz waves have uncovered the hidden message (*imaging area outlined in red*).

What Lies Underneath

Why the cross? From medieval to early modern times in Europe, lead crosses inscribed with a prayer asking for forgiveness for the deceased's sins, or with other language identifying the deceased, have been included in some burials. Such crosses have a long history, and their purposes are still a matter of debate. Did they have a theological underpinning, or was it closer to superstition? Did they originate in Christian or pagan practice? They're called, in French, *croix d'absolution* or *croix d'identité*, with similar objects found in France, Germany, and England.

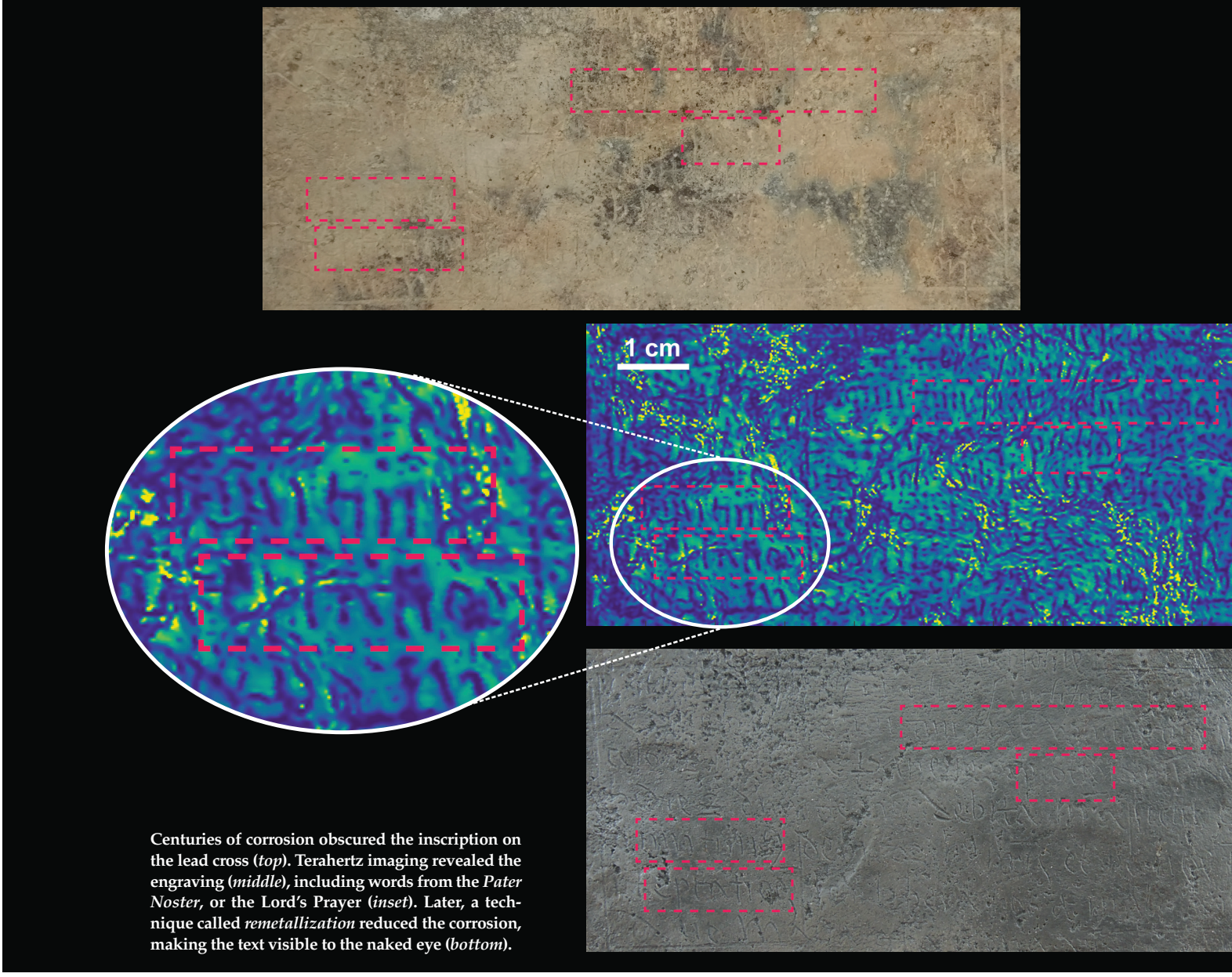
The largest collections of *croix d'absolution* have been found in Metz and Remiremont, France, in what was then the region of Lorraine and is now part of Grand Est. The cross we studied was unearthed from a burial of a high-status individual associated with the Abbey of Remiremont; a collection

QUICK TAKE

Imaging technology developed for industrial applications can also help archaeologists uncover messages on lead artifacts that are hidden under centuries of corrosion.

Terahertz electromagnetic radiation can be used for nondestructive imaging that provides a peek into the history of an artifact without causing damage.

The multidisciplinary process of developing the imaging technology demonstrates how bringing the humanities into laboratories can result in breakthroughs for all involved.



Centuries of corrosion obscured the inscription on the lead cross (*top*). Terahertz imaging revealed the engraving (*middle*), including words from the *Pater Noster*, or the Lord's Prayer (*inset*). Later, a technique called *remetallization* reduced the corrosion, making the text visible to the naked eye (*bottom*).

Junliang Dong

of crosses found there is now on exhibit in the Musée Charles de Bruyère, also in Remiremont, and other crosses are exhibited in the crypt of Cathedral Saint-Étienne in Metz. Many of the crosses are damaged and inscriptions can only be read on some of them. To interpret these remarkable objects requires gathering as much information about as many of them as is possible.

These crosses were typically made from lead, not only because it is easily smelted, cast, and worked, but because it has been recognized since ancient times to be a corrosion-resistant metal. For millennia, lead has been chosen for water pipes and for sarcophagus linings due to its ability to survive underground and in other harsh environments. *Corrosion resistant* is a relative term, however. Over the long run, lead is subject to surface and bulk corrosion, depending on environmental conditions—hence the challenge of analyzing the cross that was brought to us.

The cross we studied was first documented in 1904, shortly after its excavation from an ancient cemetery at the Abbey of Remiremont, when it was noted that its inscription was "*Très peu lisible*," that is, "very hard to read." When the cross arrived in our lab, the inscription was indeed quite unreadable. Metallic lead was hardly visible; most prominent was an irregular, multicolor top layer with a complex appearance due to a mix of corrosion products. Although the lead carbonates and oxides composing surface corrosion can be chemically reduced to metallic lead, there may also be bulk damage to the underlying lead due to a thick buildup of corrosion that is not itself reversible.

The archaeological lab asked our team to try to see whether we could read the inscription underneath the lead corrosion products. In a way, it was a test. At the time, we did not know that some of the corrosion could be chemically reversed, and indeed a collection of crosses from Remiremont would be subjected to such a treatment. Although the chemical recovery of metallic lead from the corrosion products may be successful for surface corrosion (that is, a thin, somewhat uniform corrosion layer), it often fails with bulk corrosion, in which the corrosion is not confined to a well-defined layer.



Fonds Denkmalarchiv, DRAC Grand Est, UDAP de Moselle

As seen in 1904, the Metz Cathedral Saint-Étienne, located in northeastern France, underwent an excavation of its burial sites at the turn of the 20th century. The collection of lead funerary crosses on exhibit in the crypt today was found at that time.

In many cases, then, it is not feasible to remove or reverse the corrosion; even if it were, researchers are often reluctant to use a process that alters the surface of the object. What is needed, therefore, is a nondestructive approach to seeing beneath the surface of the artifacts.

Workhorse methods that have been applied in somewhat similar situations include infrared photography, infrared thermography, x-ray imaging, and x-ray fluorescence imaging. Infrared imaging and thermography have been used for subsurface studies of paintings, for example, whereas x-ray techniques have been applied to objects as diverse as paintings and mummies.

For the lead cross, such approaches would have been unfruitful: They either cannot penetrate the corrosion, or they cannot clearly differentiate the corrosion from the underlying metal. The thick corrosion layer on the object studied would be opaque to infrared light, rendering infrared photography and infrared thermography of little use. The corrosion products of lead include, well, lead, and there is not great contrast in practice in x-ray imaging between the corrosion layer and the underlying metal. As for x-ray fluorescence imaging, both the incident x-rays and the fluorescent photons are largely attenuated by lead, making this approach difficult.

In short, among all these techniques, there were not any good, nondestructive options available for analyzing a highly corroded lead cross. That's why we turned to terahertz imaging.

Search but Don't Destroy

In dealing with art and archeological objects, one's options are restricted in many cases by a fundamental constraint: The object cannot be damaged during the investigation. This limitation may restrict what chemicals may be used, how the object can be handled, and even whether it can be contacted. The field of archaeometry applies techniques from the physical sciences to characterize art and archaeological objects using techniques both destructive and nondestructive.

Nondestructive techniques are dominated by various imaging methods across the electromagnetic spectrum, as well as ultrasonic imaging, though for some objects, the use of a fluid, often water, as a medium for the ultrasonic signals may result in damage. In the case of the lead cross, we used an imaging technique employing terahertz-frequency electromagnetic radiation, where there is no such coupling fluid or indeed contact between the terahertz equipment and the object.

The use of terahertz imaging to study art and archaeological objects is something of a side business for many terahertz research groups, although only a very small number of them work in this area full time. Terahertz techniques have been applied to paintings, drawings, ceramics, manuscripts, and mummies, to name a few examples; one application of special interest to us is measuring and mapping out the layers of paint in artwork composed of multiple layers, all without contacting or damaging the painting. But we also dabble in other types of objects—such as lead crosses.

Terahertz imaging gets its name from the type of electromagnetic waves it uses. They have frequencies ranging from 100 gigahertz to 10 terahertz, putting them broadly in the "terahertz" range; their wavelengths are typically less than a millimeter long. The imaging technique entails beaming short terahertz pulses at an object, whose different layers will reflect, or echo back, the pulse, with a time differential from layer to layer. The time measured between echoes can be used to create a thickness map of corrosion over the object, helping us to see what lies beneath. This process is then repeated at points on the sample separated by 200 microns, forming an image in the end. Terahertz imaging, which falls under the heading of "nondestructive evaluation," is currently used, for example, for automotive paint deposition monitoring, for measuring the thickness of coatings on pills, and for circuit testing. New applications continue to emerge as costs come down and terahertz imaging proves itself in new contexts.

The raw data we collect are often not enough to create a clear picture, so we have to rely on extensive processing techniques. Sure enough, when we investigated the terahertz data for the lead cross, a visual inspection showed no obvious evidence of an inscription. But that didn't mean the evidence was not buried in the data. To pull a useful image out of the information stored on our hard drive, we had to develop an advanced new type of image processing. This work was done in collaboration with Georgia Tech-Europe and the Georgia Tech-CNRS International Research Lab, along with the Musée Charles de Bruyère and the Laboratoire d'Archéologie des Métaux in Jarville-la-Malgrange. (CNRS is the French National Center for Scientific Research.)

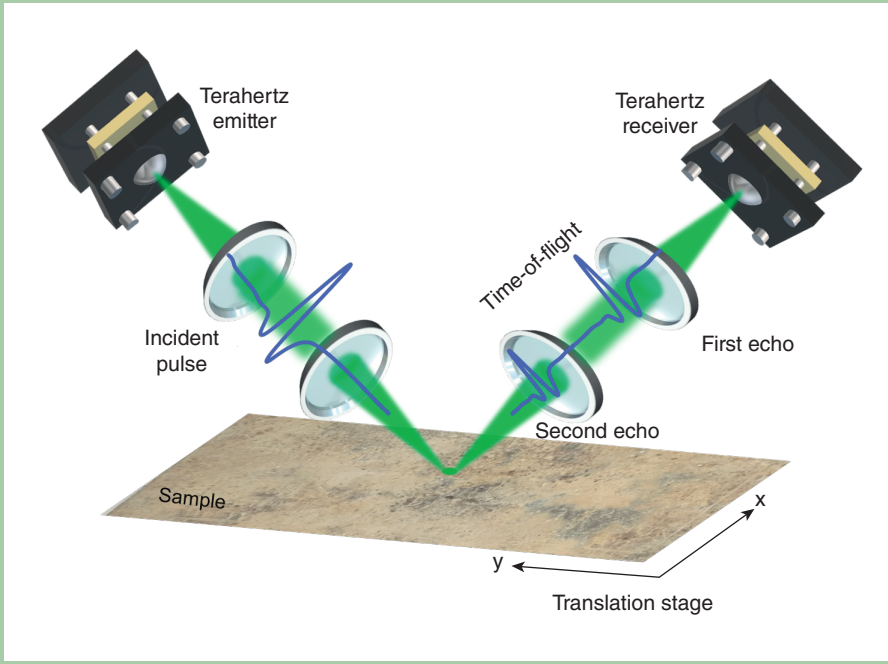
The Terahertz Gap

Terahertz waves occupy the part of the electromagnetic spectrum that lies between microwaves and infrared waves, which is referred to as the *terahertz gap.* This in-between band of wavelengths has long presented technological problems. It has been of interest for decades mainly to astrophysicists and a small community of experimentalists who have developed specialized molecular lasers operating at terahertz frequencies. Before a major advance in the mid-1980s, the existing technologies covering the terahertz region were expensive and lacked tuning and flexibility. The technology became more widely available with the explosive growth of ultrafast optical technology.

In 1985, pioneering physicists Donna Strickland and Gérard Mourou published an influential paper that described a practical way to generate sub-100-femtosecond optical pulses (pulses that last less than 10 trillionths of a second). From this point on, ultrafast pulses—typically in the near infrared—were available in numerous labs on the tabletop. In addition to winning the researchers the Nobel Prize in Physics (in 2018), their work opened a path to the generation of terahertz waves in the lab, which began in the late 1980s.

Ultrafast optical pulses can be used to seed the terahertz pulses using a technology known as an Auston or photoconductive switch. When a subpicosecond pulse of near-infrared laser light impinges on the Auston switch, it triggers a transient current that acts as the source of a pulse of terahertz electromagnetic radiation. The result can be a transient that looks like a single electromagnetic cycle. By modulating picosecond-scale infrared laser pulses, it's possible to generate a spectrum of terahertz waves ranging in frequency from 100 gigahertz to a few terahertz.

Early pulsed terahertz systems were costly and cumbersome. Today, such systems are commercially produced at far lower cost, making the technology more accessible. In our lab, we commonly use the systems to study fiber composites and plastics, and to measure paint-layer thicknesses. Terahertz waves can be used to image in three dimensions the interior of a range of electrically insulating materials, providing a technique to sense and, even more, to locate defects that were incorporated into components in manufacturing or have developed in service. As such, terahertz imaging provides a technique to carry out nondestructive evaluation for several industrial applications.



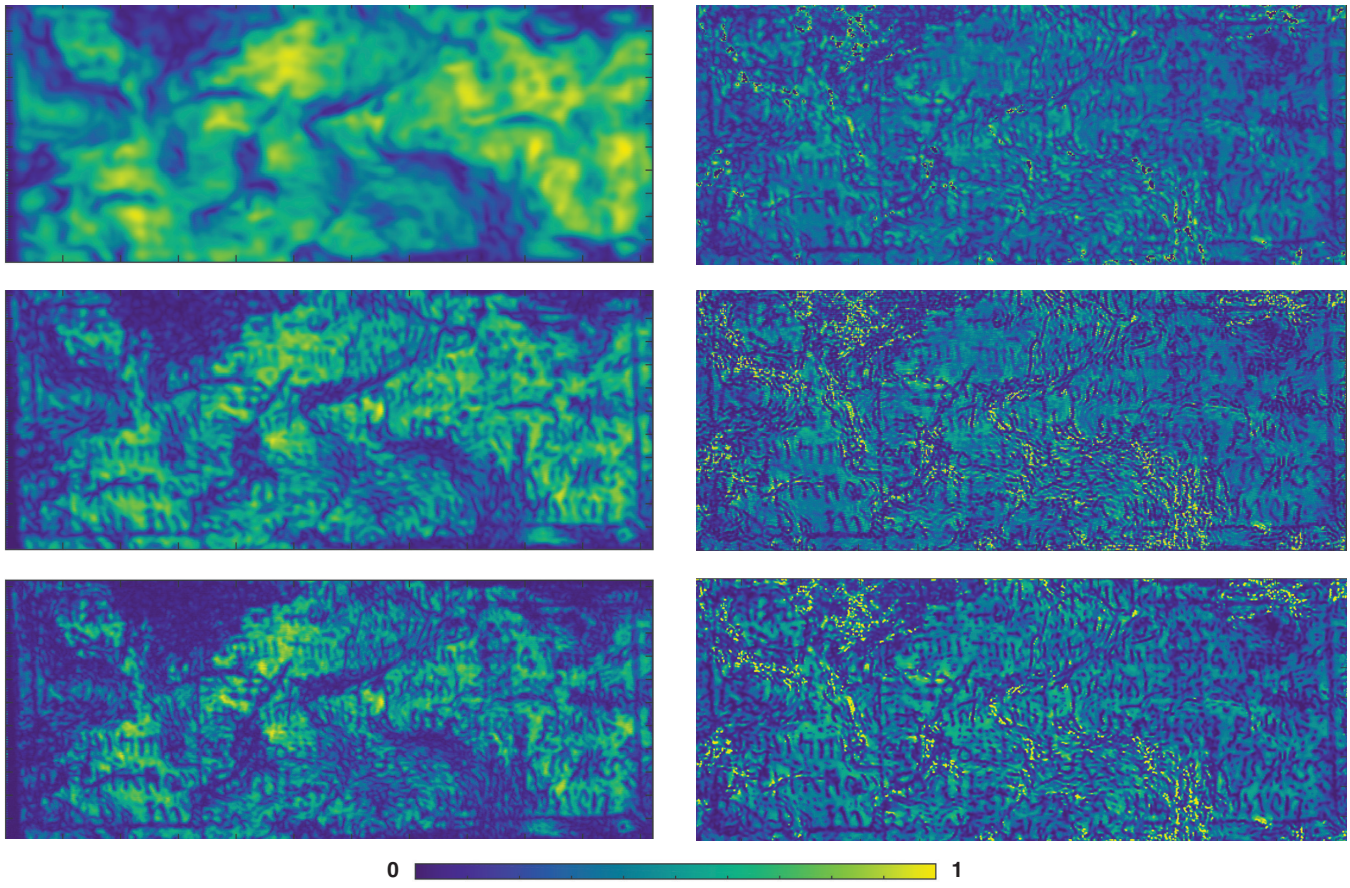
Junliang Dong

To produce an image using the *reflection technique*, pulses of terahertz waves are sent from an emitter to reflect off the cross—from both the surface level and the deeper level of lead—back to the receiver. The time differences between the reflections are measured to create a 3D map of the corrosion.

For the present study, we worked in reflection, focusing terahertz pulses on the sample and measuring the reflected pulses in time. This technique maps out the temporal shape of the reflected pulses using a photoconductive terahertz receiver that operates somewhat like the emitter, but in reverse. We would move to a neighboring pixel on the sample using a translation stage, rastering across the lead cross in two dimensions in 0.2-millimeter steps. After the raster scans, a 3D volume of reflected data had been acquired (based on the two-dimensional plane of the cross, plus the dimension of time to gauge depth).

Using this method, the reflected terahertz pulses provide both surface and subsurface information across the lead cross. The variations in the reflections correspond to irregularities or discontinuities in the structure at the air/corrosion and corrosion/lead interfaces, and they contain information related to the sought-after inscriptions.

We expect that there will be reflected pulses or echoes from two layers: the surface of the corrosion and the lead underneath. However, the two reflections are not visually evident in the raw reflected signal because of the thinness of the corrosion layer, so we use a deconvolution algorithm (or what's called *frequency-wavelet domain deconvolution*) to resolve the two overlapping echoes. The first, strong echo seen is from the corrosion's surface layer; the second, weaker echo is from the lead beneath the corrosion. By measuring the time between these two echoes, we create a thickness map that reveals a highly irregular corrosion-layer thickness, ranging from tens to hundreds of microns. This map allows us to zero in on the underlying lead layer to find information about the hidden inscription.



Junliang Dong

The authors created these terahertz images through a two-stage imaging process. First, multispectral images at increasing terahertz frequencies (*scale bar shows 0 to 1 terahertz*) began to uncover the text (*left column, increasing from top to bottom*). Then, post-processing methods including intensity correction (*top right*), the use of a deconvolution algorithm (*middle right*), and contrast enhancement (*bottom right*) further revealed the inscription.

Just the Image, Please

The starting point for our image processing was to Fourier transform the electric field of the terahertz pulses produced by our apparatus—a data-processing step that identifies the frequencies present in the terahertz pulses in the form of a spectrum. We then sliced the spectrum into six frequency bands, each with a bandwidth of 0.5 terahertz. The reason for doing so was that these various frequency bands contained different information about the corrosion/lead interface, in part because higher-frequency bands contained information about finer structure. We could use these differences to distinguish the inscription. In a very rough sense, some frequency bands said more about the damage than the underlying object, so by factoring out undesired information, we hoped to be left with a clear image of the text.

The terahertz multispectral images were obtained by taking the Fourier transform of the reflected terahertz signal at each "pixel" of the object, that is, at each point where we took a measurement, and integrating the strength of the spectrum within the various bands. Based on these images, we were able to see the inscription emerging in the high-frequency bands. Still, the inscription in the images associated with these high-frequency bands was obscured by corrosion. Moreover, the image contrast was poor due to the relatively low spatial resolution provided by the terahertz measurement.

In order to overcome these limitations and develop a clearer image of the cross beneath its corrosion layers, we developed a post-processing chain for digital image restoration and enhancement to improve the legibility of the inscriptions. Such a post-processing strategy allowed us to enhance the features associated with the inscriptions observed in the high-frequency bands, while removing undesirable remnants based on the information provided in the low-frequency bands.

Our image-processing chain contained four steps. First, we corrected for variations in gross surface morphology of the corroded lead cross that had a substantial impact on the terahertz reflections. These included contributions from damage due to bulk corrosion as well as to presumed hammer marks that were left in working the lead sheet from which the cross was cut. We considered the low-frequency image as a good estimate of the intensity variations due to surface morphology, in which the influence of the inscriptions was excluded. Then we used this estimate to correct the intensity variations in the high-frequency image, where the inscriptions existed.

In our second step, we worked to fill in dark areas in the image that contained little evident detail. We implemented an image inpainting algorithm, which successfully filled in the missing areas based on the image formation available outside the dark patches. Third, we removed the image blur induced by the limited spot size of the terahertz focus on the sample, in turn leading to a higher-spatial-resolution image.

Fourth and lastly, we enhanced the image contrast by directly setting the

intensity of the areas considered as inscriptions to zero.

The final image we achieved after post-processing indeed reveals a significant portion with features associated with inscriptions. Although the text is recoverable because of its formulaic nature, some areas of the inscription remain obscure due to bulk damage from corrosion—rather than just from a surface layer corrosion—which resulted in some dark areas and cracklike features in the final image.

Heaven Again

By this point, a substantial portion of the inscription on the cross had become readable. One of the collaborators on the project, Aurélien Vacharet of the Musée Charles de Bruyère, is an art historian, and he had no trouble at this stage completing the puzzle and reading the Latin text incised in the lead in cursive Carolingian minuscule: It was the *Pater Noster* (also known as The Lord's Prayer). The *Pater Noster* is a fairly typical prayer in this context, so the discovery of the text was not Earth-shattering, but to read the text through the corrosion was a breakthrough.

Following the completion of our project, the cross then left our lab for the Laboratoire d'Archéologie des Métaux, where Ana Ribeiro was responsible for electrochemically reducing the corrosion, remetallizing the corrosion to lead. Although that remarkably revealed the inscription to the naked eye, there was still damage evident, most notably in the heavily pitted areas. These were precisely those regions where the image processing did not yield good results. It is worth noting, as well, that the hammer marks on the lead also showed up in the processed images. Indeed, the entire Remiremont collection underwent this process with excellent results overall; however, a number of crosses have been so heavily damaged over the years that little in terms of inscription could be recovered.

So why were we approached? In a sense, it was about more than simply the individual cross. In the case of the cross we studied, the reversal of the corrosion was fairly successful, but in many cases it was not. It is therefore advantageous to have a bag of tricks available for situations when the success of corrosion reversal is in doubt, or the process poses too much of a risk to the object. So our experiment was a proof of principle that we hope will be dusted off when the need arises.

Earlier, we referred to such objects as *croix d'absolution*, which the lead cross seemed likely to be when it was first brought to us; however, the *Pater Noster* is not a prayer of absolution. As it turned out, none of the Remiremont crosses were inscribed with obvious absolution prayers, but other crosses in the collection were inscribed with the *Pater Noster* as well.

There is still work to be done on the historical context of these crosses—part of a process of discovery that speaks to many questions about the past, touching on subjects including attitudes toward death, how such attitudes varied according to geography and time period, and their relationship with social class. Of course, any interpretation of inscriptions of such crosses must also account for the date of the object and where it was produced and found. So, regardless of the origin of the practice of including such crosses in burials, by early modern times this practice may have lost its original meaning and more or less become simply a propitious custom.

To pull a useful image out of the information stored on our hard drive, we had to develop an advanced new type of image processing.

A Multidisciplinary Journey

Our team was fortunate that this particular cross ended up in our lab. It was in bad enough condition to provide a challenge for terahertz imaging, but not so bad that the challenge defeated us. The technique we honed can now be applied to various other historical lead objects, including ones that might not be candidates for remetallization (possibly due to fragility, size, or being in situ with other objects). The knowledge that might be gained through this process is considerable, given lead's widespread use in the ancient world.

Artifacts composed of other types of metals might also be candidates for terahertz analysis. For example, we have recently produced a preliminary analysis of a corroded bronze Byzantine coin and have imaged the profile of the emperor underneath the corrosion. Already, terahertz imaging has attracted interest in using it to study paintings, ceramics, plastics, textiles, manuscripts, and wood and stone objects.

If there is a larger lesson here, maybe it is how to bring the humanities, science, and engineering into the same room. There is a network of scientists and engineers, laboratories, and conservation scientists working on understanding and preserving our cultural histories. This effort is, in our view, certainly a worthy pursuit for our field. On a personal level, collaborating with art historians, medieval scholars, and archaeologists to read a previously hidden, inscribed message was richly rewarding, opening our eyes to other ways of viewing our material heritage.

Bibliography

Auston, D. H., K. P. Cheung, and P. R. Smith. 1984. Picosecond photoconducting Hertzian dipoles. *Applied Physics Letters* 45:284–286.

Chevreux, P. 1905. Les Croix de plomb dites croix d'absolution de la région vosgienne. *Bulletin archéologique du Comité des travaux historiques et scientifiques*. Paris: Impr. Nationale.

Dabrowska, E. 2005. Passeport pour l'au-delà. Essai sur la mentalité médiévale. *Le Moyen Age* CXI:313.

Dong, J., A. Locquet, and D. S. Citrin. 2015. Enhanced terahertz imaging of small forced delamination in woven glass fibre-reinforced composites with wavelet denoising. *Journal of Infrared, Millimeter, and Terahertz Waves* 37:289–301.

Dong, J., A. Ribeiro, A. Vacheret, A. Locquet, and D. S. Citrin. 2022. Revealing inscriptions obscured by time on an early-modern lead funerary cross using terahertz multispectral imaging. *Scientific Reports* 12:3429.

Hedjam, R., and M. Cheriet. 2013. Historical document image restoration using multispectral imaging system. *Pattern Recognition* 46:2297–2312.

Strickland, D., and G. Mourou. 1985. Compression of amplified chirped optical pulses. *Optics Communications* 55:447–449.

David S. Citrin is a professor in the School of Electrical and Computer Engineering at Georgia Tech. Alexandre Locquet is an adjunct assistant professor in the School of Electrical and Computer Engineering at Georgia Tech. Junliang Dong is a researcher at the Institut national de la recherche scientifique, Centre Énergie Matériaux Télécommunications in Canada. Email for Citrin: david.citrin@ece.gatech.edu

Genomic Prediction in the Big Data Era

A simple model from the early 20th century remains our best tool for using DNA to predict disease risk and other complex traits.

Gustavo de los Campos and Daniel Gianola

In recent years, more than 30 million genomes have been genotyped by companies such as 23andMe or MyMedLab. These companies provide customers with ancestry and health-related information, and link genotype data with measurements (called *phenotypes*) collected from surveys, wearables, and electronic health records. The resulting datasets are routinely used for genomics research. Likewise, several public initiatives have developed large biomedical datasets, such as the UK Biobank and the All of Us program, comprising DNA and extensive phenotypic attributes from hundreds of thousands of participants.

The advent of big data in genomics research has revealed many insights about how variation at the DNA level influences phenotypes and disease risk. It has ushered in unprecedented improvements in genomic prediction models. However, we are far from reaching the potential that these technologies can offer. A simple linear model conceived in the early 20th century remains one of our best tools and the basis for the prediction of phenotypes and disease risk from DNA. Are we primarily limited by the volume, diversity, and quality of data available, or by the techniques used to learn from that data? Will machine learning further advance the prediction of traits and disease risk using DNA information? These questions are crucial for understanding how (and why) currently used genomic prediction methods work, what the main barriers limiting our ability to predict disease risk from DNA are, and how such barriers can be overcome.

From Mendel to Polygenic Models

The current understanding of hereditary genetics is based mainly on Gregor Mendel's laws of inheritance. Using data he collected in hybridization experiments, Mendel proposed a model that, in its simplest form, consisted of one genetic locus ("site") with two discrete segregating "particles" (later called *alleles*) that could explain the patterns of inheritance that he observed in his experiments. It took a long time after that for scientists to discover the existence of DNA and genes. Since then, advances in molecular biology have uncovered myriad details about the recombination of parental alleles during meiosis and the intricate ways that genetic and environmental factors shape gene expression. Yet Mendel's simple model captures the essence of inheritance.

Mendel's laws describe the inheritance of qualitative traits such as the color and texture of seeds. For decades thereafter, scientists attempted to develop mathematical models to describe the inheritance of quantitative traits such as human stature. In 1918, the influential statistician R. A. Fisher proposed a simple linear model that posed how Mendel's theory of inheritance could explain the descent of quantitative traits such as human stature or body weight. We refer to this model as the *Fisher–Wright infinitesimal* model, because Sewall Wright developed an equivalent model.

A Simple and Robust Model

Molecular biology suggests that genes and environmental factors affect phenotypes in complex (nonadditive, nonlinear) ways, including recursive and feedback effects, gene-by-gene interactions, and gene-by-environment interactions. The Fisher–Wright infinitesimal model (*see the sidebar on page 288*) ignores gene-by-gene and gene-by-environment interactions and many other complexities of the biology underlying most human phenotypes and diseases. But its simplicity and robustness have proved to be very effective at describing and predicting patterns of variation and inheritance for most traits. The Fisher–Wright model has been widely adopted in animal and plant breeding for predicting the traits of candidates of selection—for example, the expected milk production of the daughters of a young bull. More recently, this approach has been adopted to predict human traits and disease

QUICK TAKE

In 1918, R. A. Fisher proposed a model for hereditary genetics based on Gregor Mendel's laws of inheritance. It became the basis for predicting phenotypes as the field of genetics developed.

Although this model does not accurately reflect causal relationships between genotypes and phenotypes, it can predict phenotypes well and is widely used in agriculture and health care.

The era of big data in genomics has improved the accuracy of DNA-based model predictions of phenotypes. Still, the potential that these technologies offer has not yet been reached.



Wikimedia Commons/Friedrich Haag

Complex traits—such as the differences in size, color, flavor, ripening time, and texture of these tomatoes—are influenced by genetic and environmental factors. These factors can interact in nonlinear and nonadditive ways, yet for many traits R. A. Fisher's linear, additive model has proved to be very effective at describing and predicting such variation.

risks—such as the risk of developing a specific type of cancer—in individuals who are not yet expressing those traits or diseases.

Scientists have considered several explanations for why a model that neglects fundamental aspects of biology effectively predicts variation in phenotypes attributable to interindividual differences in DNA. The most accepted explanation involves a combination of two factors. First, genomes are made of discrete bits of information: alleles. Second, the distribution of allele frequencies over loci is typically skewed, making a few genotypes common and most combinations of alleles rare. This skew leads to landscapes (a map from genomes to genetic values) that can be approximated well by linear functions: straight lines for one locus, planes for two loci, and hyperplanes for multiple loci.

It is worth noting, however, that a linear approximation to a highly nonlinear function works well only within a specific domain (such as a homogeneous population). Furthermore, the specific linear model that would approximate the map from genotypes to phenotypes is expected to vary between populations.

Fisher's Additive Model in Genomics

In the 20th century, biologists employed phenotype and pedigree data as inputs in Fisher's model. More recently, modern sequencing technologies have made it possible to "read" genomes at hundreds of thousands of *SNPs* (*single nucleotide polymorphisms*, meaning locations on the genome where base pairs differ and that are commonly used to study genetic variation). This technological development enabled scientists to map loci associated with phenotypes, and it led to the implementation of an instrumental version of Fisher's additive model using SNP marker information in lieu of the "causal" loci represented in the Fisher–Wright infinitesimal model equation. Predictions from such models are known as *polygenic scores* in human genetics and *predicted genomic breeding values* in agriculture.

Modern genotyping arrays contain more than one million SNPs. Further, using whole-genome sequencing or imputation methods can produce genotypes for several million SNPs. However, for any given trait, most of these SNPs have no predictive value by themselves. Therefore, in human genetic studies, the first step in building a prediction model is to select, among the millions of SNPs available, hundreds or thousands that are associated with a particular trait or disease. This analysis is known as a genome-wide association study and consists of (statistically) testing the association of a phenotype with a SNP, one SNP at a time. Each association test produces a *p*-value, which measures the strength of statistical evidence against the absence of association (the null hypothesis in this case). A very small *p*-value provides evidence of a significant phenotype–SNP association. The

Fisher–Wright Infinitesimal Model

Fisher showed in 1918 how Mendel's laws of inheritance, originally developed for categorical traits, could also explain the inheritance of quantitative traits. Human genomes comprise approximately 3 billion sites (*loci*), each holding a pair of nucleotides, one inherited from each parent. Many of these positions are preserved, meaning that all individuals carry two copies of the same genetic code. However, there are hundreds of millions of positions in the human genome for which at least two segregating nucleotides (also called *variants* or *alleles*) have been discovered; such positions are referred to as *SNPs* (for *single nucleotide polymorphisms*).

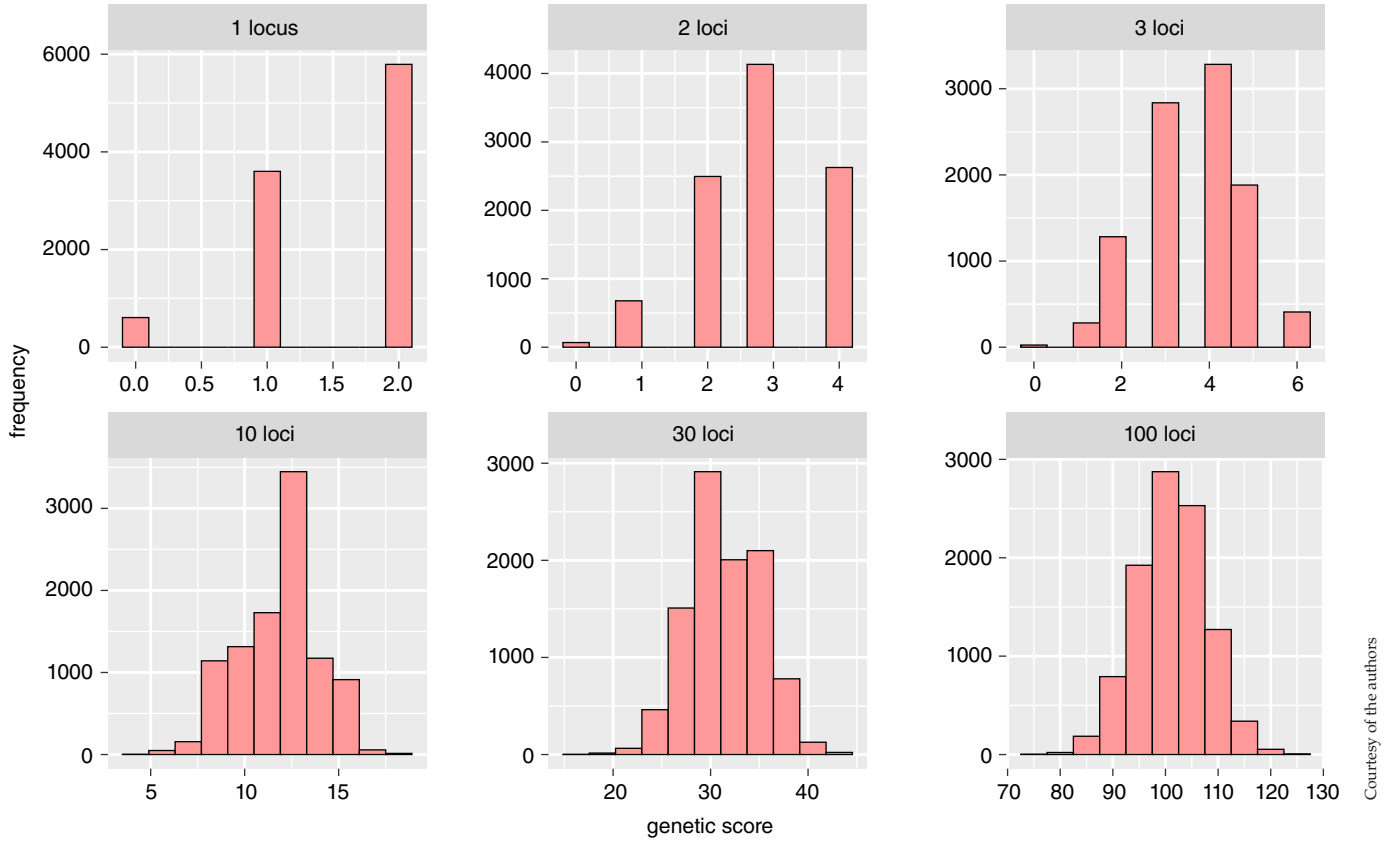
To illustrate, consider a locus with two variants, alleles A and G. There will be three genotypes in the population: AA, AG, and GG. To formulate a mathematical model for this locus, one can numerically represent an individual's genotype by counting the number of copies of one of the alleles at each of the loci that influence a trait. In this instance, count the number of copies of the G allele, and represent genotypes as 0 (AA), 1, (AG), and 2 (GG).

The Fisher–Wright additive model describes a quantitative phenotype (y_i, such as the body weight of an individual) as the sum of the effects of many variants, plus an environmental effect (reflecting the aggregate impact of many environmental factors). In its simplest formulation, the model takes the form of the regression:

$$y_i = \mu + Z_{i1}\alpha_1 + Z_{i2}\alpha_2 + \ldots + Z_{ip}\alpha_p + \varepsilon_i$$

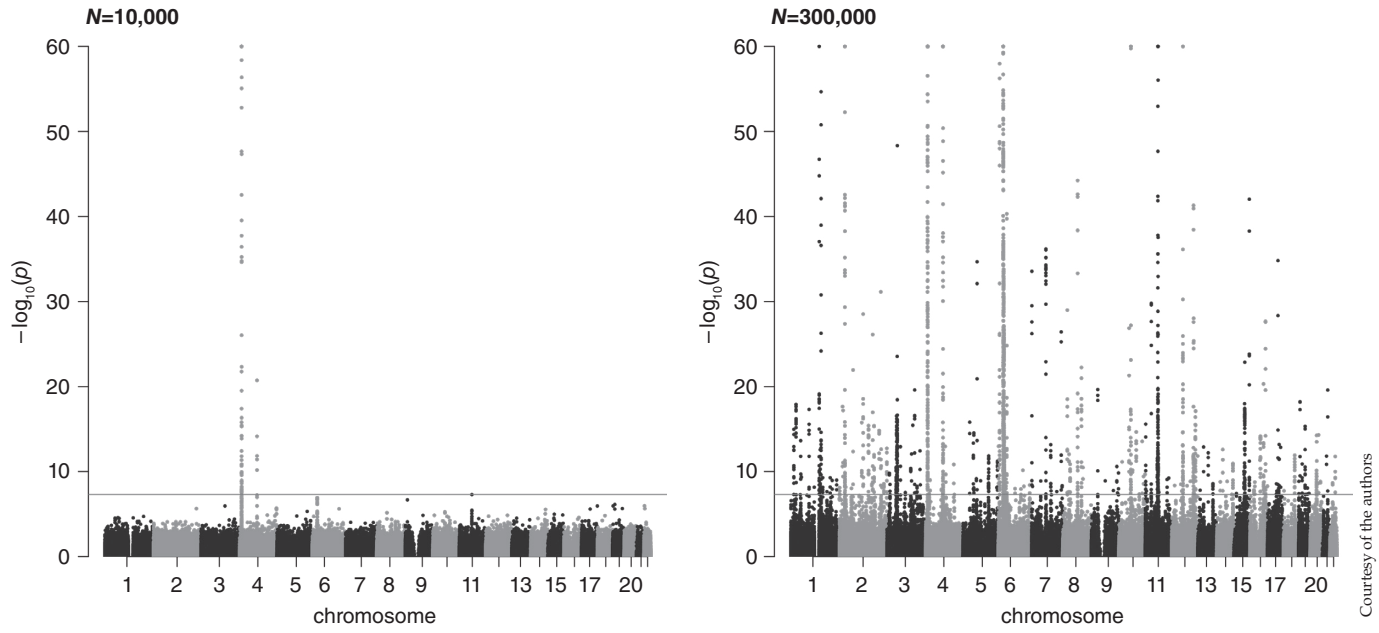
Above, μ is a constant; Z_{ij} is a numerical representation (for example, 0, 1, or 2) for the genotype of the ith individual at the jth causal locus ($j = 1, \ldots, p$); α_j is the additive effect of a reference allele at the jth locus; and ε_i is a residual term capturing environmental effects, and possibly other genetic effects not captured by the linear regression. The Fisher–Wright model consists of the limiting situation where the number of loci is infinite; for this reason, the model is also referred to as the additive infinitesimal model. However, the model provides a reasonable approximation when the number of loci is moderately large.

For a single locus, the regression function of this linear model generates three discrete values in the population (μ, $\mu + \alpha_1$, and $\mu + 2\alpha_1$) corresponding to individuals who carry none, one, or two copies of the reference allele, respectively. However, for situations involving many loci, the sum of effects across loci ($\mu + Z_{i1}\alpha_1 + Z_{i2}\alpha_2 + \ldots + Z_{ip}\alpha_p$) produces a seemingly continuous, bell-shaped distribution of genetic values (*see the figures below*).



Courtesy of the authors

When the genetic basis of a trait involves just a few loci (*as shown in the top panels*), the distribution of the genetic scores shows clear, discrete variation; however, as more loci are involved, the scores fill in the gaps and converge to a normal distribution (*as shown in the bottom panels*). This simple phenomenon can be used to explain how the aggregation of effects from many loci, each of which aggregates in a Mendelian (discrete) fashion, forms the basis for the inheritance of quantitative traits.



Courtesy of the authors

As dataset size (*N*) increases, more SNPs and chromosome segments are found to be associated with uric acid levels (serum urate). In the Manhattan plots above, the evidence of association of individual SNPs with serum urate is displayed by chromosome and position. The higher the plot scores on the *y* axis, the lower the *p*-value of the association. The horizontal line represents the threshold of significance; points above that line are significantly associated with serum urate. The table on the right shows the total number of significantly associated SNPs and the number of segments of 100 kilobase pairs containing at least one significantly associated SNP.

sample size	number of significant results	
	segments of 100 kilobase pairs	individual SNPs
5,000	2	37
10,000	3	66
30,000	10	235
50,000	20	357
100,000	58	758
300,000	210	2,530

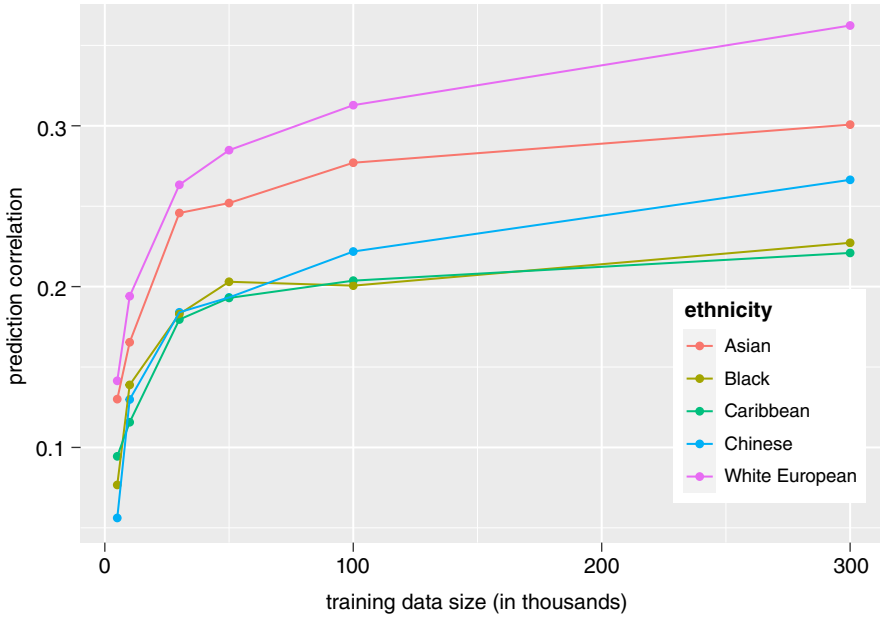
results of a genome-wide association study are displayed in a *Manhattan plot*, a name that refers to its resemblance to the New York City skyline. (*See the figure above, where we provide the result of a genome-wide association study in which we tested the association of around one million SNPs with serum urate, which is a risk factor for gout.*)

The GWAS (Genome-Wide Association Studies) Catalog (www.ebi.ac.uk/gwas/) compiles results published in more than 45,000 studies, collectively reporting close to half a million loci associated with one or more of 5,000-plus phenotypes included in the catalog. It represents a stunning advance in only two decades of genome-wide association studies. These studies have spawned research in many fields. For example, genome-wide association studies have identified many large-effect variants associated with sizable increases in the risk of developing breast and ovarian cancers, and some of this information is now used to screen for patients at high risk of developing these cancers.

However, most traits and diseases are influenced by many small-effect variants, with no single variant providing high predictive power. For such phenotypes, prediction is made by accumulating the risk conferred by many small-effect variants, using a calculation called a *polygenic score*. Many of the direct-to-consumer genotyping companies provide clients with the predicted risk of some conditions. Genome-wide association studies have also revealed numerous chromosome regions that are good candidates for pharmacogenetic studies.

The advent of big data in genomics research has improved our ability to map variants and predict complex traits at an unprecedented pace.

The power to detect an association between a SNP and a phenotype (or disease risk) depends mainly on the proportion of variance of the phenotype that can be "explained" by the SNP and the size of the dataset used to test the association. Typically, individual SNPs explain a very small fraction (considerably less than 1 percent) of the variance of a trait. Therefore, to achieve high power in detecting association, genome-wide association studies need to use a very large sample size. We illustrate this concept in the figures above, which show results from our genome-wide association analysis for serum urate, using data from the UK Biobank. In a sample size of 10,000,



Prediction correlations of polygenic scores derived using significant SNPs—here from an analysis of the data from the figures on page 289—vary based on the training and testing datasets. In this case, the training data were from white Europeans. The accuracy of the polygenic score is compared between testing datasets from different ancestry groups (all from the UK Biobank).

only three segments of 100 kilobase pairs harbored SNPs significantly associated with serum urate; however, with a sample size of 300,000, 210 segments had at least one significant SNP.

Prediction in the Big Data Era

The first wave of genome-wide association studies, starting about two decades ago, relied on cohorts of a few thousand individuals. However, it soon became clear that larger sample sizes are needed to detect variants that are weakly associated with phenotypes. Subsequently, efforts were made to attain larger sample sizes, first by meta-analyzing summary statistics from multiple cohorts and, more recently, using biobank-size datasets in the public and private domains. The advent of big data in genomics research has improved our ability to map variants and predict complex traits at an unprecedented pace. A large sample size also improves the accuracy of effect estimates (the β's in the equation on page 292). These two improvements have heightened the precision and accuracy of polygenic scores.

The impact of big data on polygenic score prediction accuracy is illustrated in the figure on the left, showing the prediction correlation of a polygenic score for serum urate (derived with data from Europeans) in testing data by the size of the dataset used to derive the polygenic score (meaning to discover significant SNPs and to estimate SNP effects). The top curve in that figure shows the prediction correlation achieved when the testing data were also from Europeans. (In other words, the testing data and training data were from the same population.) The correlation between the polygenic score and serum urate increases with sample size, achieving a prediction correlation of 0.37 with a sample size of 300,000.

Achieving this moderately high prediction correlation (nowadays standard for many traits and diseases) is remarkable, considering that serum urate is highly affected by many other factors, notably diet and lifestyle. The trait heritability bounds the prediction accuracy of a polygenic score. However, owing to limitations of the data available (notably the sample size) and the methods used, most polygenic scores achieve correlations smaller than the upper bound imposed by the trait heritability. This implies that polygenic scores are not a tool for diagnostics; however, DNA-based predictions can be effective for screening for disease risk.

Interestingly, in the figure on the upper left, the curve relating the prediction correlation of the polygenic score with sample size does not reach a plateau with a sample size of 300,000, suggesting that larger sample sizes would render even higher prediction correlations.

The Additive Model's Influence

The Fisher–Wright infinitesimal model has profoundly affected science. In statistics, Fisher's 1918 publication introducing this additive model also provided the basis for the analysis of variance (ANOVA), a technique commonly used in data analysis. In ANOVA, the dispersion of an outcome (y_i) is partitioned into components. For instance, we can rewrite the equation in the first sidebar (at the top of page 288) in terms of three components:

$$y_i = \mu + g_i + \varepsilon_i$$

where μ is a constant (meaning it does not vary between individuals), and $g_i = Z_{i1}\alpha_1 + Z_{i2}\alpha_2 + \ldots + Z_{ip}\alpha_p$ and ε_i are subject-specific terms that capture the contribution of genetic and nongenetic factors to interindividual differences in phenotypes, respectively.

In genetics, the proportion of interindividual differences in phenotypes "attributable" to genetic factors is known as *heritability*. This parameter plays a central role in quantitative genetic theory: It relates to the resemblance between relatives, to the expected response to selection, and to the "accuracy" of genetic prediction.

The additive model predicts that the correlation between the phenotypes of two related individuals (that do not share the same environment) is proportional to the proportion of alleles shared by descent, times the trait heritability.

Because the expected proportion of alleles shared by relatives can be derived from pedigree, Fisher's model can be used to estimate genetic parameters and to predict genetic values using phenotype data and pedigree information (without DNA data). Today, predictions from additive models are used routinely to breed for desirable phenotypes both in animal and plant breeding.

The success of the additive model in the genomic era has been remarkable. However, the model and the data used to train it have limitations worth considering.

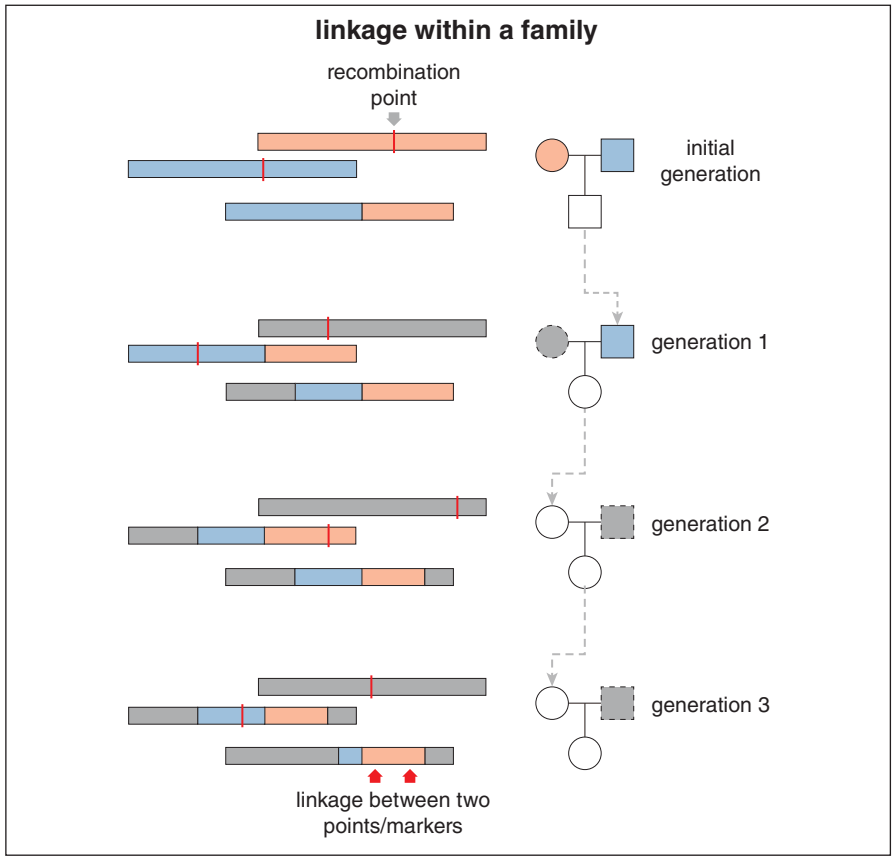
Linkage Disequilibrium Challenges

Genotypes at SNPs that are physically proximal in the genome tend to be correlated. This phenomenon, called *linkage disequilibrium*, makes it possible to map causal loci and develop polygenic scores that predict phenotypes even when causal variants have not been genotyped. Linkage disequilibrium can also occur when alleles at SNPs in different chromosomes are correlated due to selection pressure on some favorable combinations (called *epistasis*, a form of nonadditive gene function). The fact that we use models based on imperfect surrogates imposes limitations on genomic prediction models. For example, the use of SNPs that are imperfectly correlated with causal loci makes the upper bound on the prediction accuracy of a polygenic score considerably lower than its biological upper bound.

Another imperfection introduced by linkage disequilibrium is that genomic models may not provide unbiased estimates of genetic parameters, irrespective of the method of estimation used. For instance, SNP-based models typically don't produce unbiased estimates of the trait heritability or the genetic correlation between traits simply because such models are based on SNPs that imperfectly tag causal alleles.

The use of models based on SNPs that are imperfectly correlated with alleles at causal loci also affects the portability of polygenic scores between populations. (In the figure on the facing page, the prediction correlations for polygenic scores for serum urate derived with data from Europeans were also applied to groups of non-European ancestry.) Finally, the use of SNPs that are imperfect surrogates of the genotypes at causal loci makes the study of gene–gene interactions challenging, a phenomenon that we call *phantom epistasis*. In such a situation, a system that does not involve gene–gene interactions at the causal level may lead to statistically significant SNP–SNP interactions simply because the SNPs used in the analysis do not perfectly tag the genotypes at causal loci.

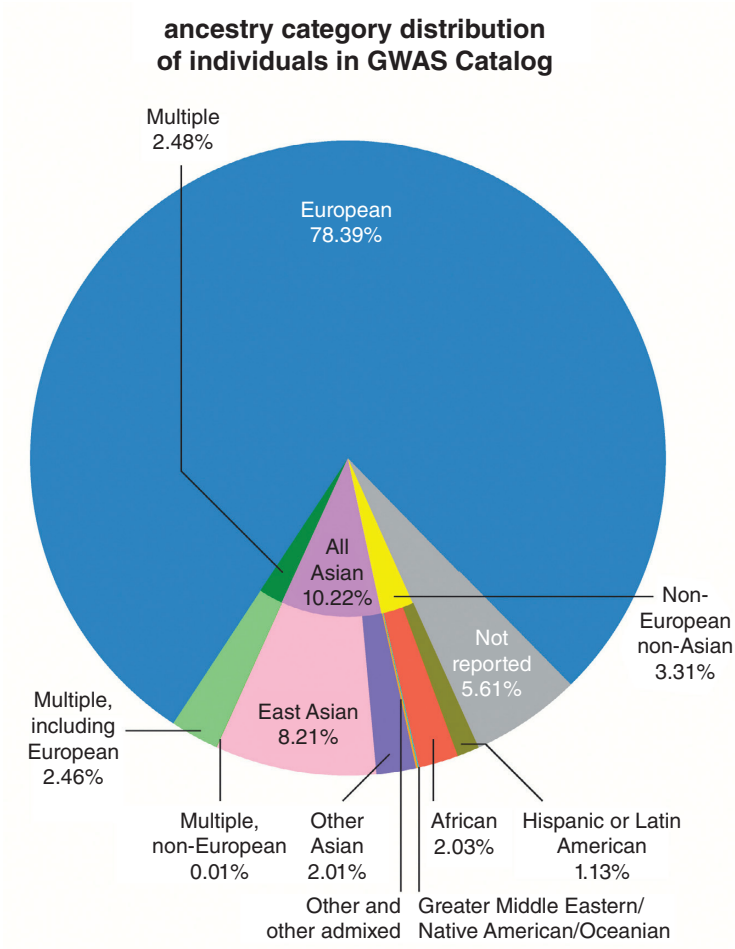
Modern imputation techniques can predict genotypes at millions of SNPs using the SNPs included in the genotyping array (around one million SNPs) and reference panels of whole-genome sequenced samples. The ultrahigh SNP density produced by imputation increases the probability of having SNPs in high linkage disequilibrium with causal variants, which may improve prediction performance and the portability of polygenic scores between populations. However, although imputation can slightly improve the accuracy of a polygenic score, the increase in prediction correlation is typically more limited by the sample size than the number of SNPs available in the array. More recently, as the cost of whole-genome sequencing continues to decrease, many initiatives (such as the All of Us program, a U.S. National Institutes of Health research program that is building a diverse database of genotypes and diseases) are now providing SNPs derived from whole-genome sequencing. Conceptually, whole-genome sequences contain all the causal variants, which could improve polygenic score prediction accuracy. However, as noted previously, the sample size may limit the increase in prediction accuracy achieved by using whole-genome sequencing.



Wikimedia Commons/William S. Bush, Jason H. Moore

Genotypes that are physically proximal in the genome tend to be correlated. At the population level, this phenomenon—called *linkage disequilibrium*—makes it possible to map causal loci and develop polygenic scores that predict phenotypes even when causal variants have not been genotyped.

Machine learning methods do not systematically outperform additive models for genomic prediction—and often underperform.



From Sirugo, Williams, and Tishkoff, 2019.

Individuals of European descent account for about 16 percent of the global population but represent almost 80 percent of all participants included in genome-wide association study (GWAS) datasets. Many of the findings in such studies do not replicate in populations of non-Europeans. Several initiatives are working to expand diversity and representation in genome-wide association datasets.

Non-Europeans Are Underrepresented

Individuals of European descent account for approximately 16 percent of the global population but represent almost 80 percent of all participants included in genome-wide association datasets. Many of the findings in studies that include only Europeans do not replicate in populations of non-Europeans. Likewise, the polygenic scores developed in one population do not perform very well in cross-ancestry prediction (*see the figure on page 290*). Scientists are aware of the poor portability of polygenic scores across populations and how that contributes to research disparities that may lead to increased health gaps. There are some efforts to solve this problem by creating more representative datasets (such as the All of Us program) as well as developing new methods that can lead to better polygenic scores for cross-ancestry prediction.

Multiple factors can make the genetic architecture of a trait population-specific. For example, some alleles may be fixed or nearly so in a popula-

Polygenic Score Prediction

In the early 21st century, agricultural scientists studying animal breeding came up with the idea of predicting phenotypes using large numbers of SNPs distributed over the entire genome, and the method rapidly expanded into plant breeding as well. In human genetics, findings reported in genome-wide association studies are increasingly used to develop *polygenic scores*, which aggregate the effects of many variants into a single score that predicts a phenotype or disease risk. To develop a polygenic score, scientists usually fit a linear model, such as the equation on page 288 from the Fisher–Wright model, to SNPs that have cleared some significance threshold of association with a phenotype of interest in a genome-wide association analysis. The model is,

$$y_i = \mu + X_{i1}\beta_1 + X_{i2}\beta_2 + \ldots + X_{iq}\beta_q + \varepsilon_i$$

and is similar to the Fisher–Wright additive model. Using the previous equation, a polygenic score is calculated by taking the weighted sum of the genotypes at the selected SNPs:

$$PGS_i = \hat{\mu} + X_{i1}\hat{\beta}_1 + X_{i2}\hat{\beta}_2 + \ldots + X_{iq}\hat{\beta}_q$$

Above, PGS_i is the polygenic score value of the ith individual, and $\{\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_q\}$ are SNP effect estimates that act as weights for each of the SNPs in the polygenic score.

Conceptually, Fisher's earlier model (the equation on page 288) includes as predictors the individual's genotypes at all loci affecting the trait (the Z variables on the right-hand side of the equation on page 288). However, the equation above is based on SNPs selected in a genome-wide association study. Genotyping arrays cover just a fraction of the genome; thus, many of the causal loci may be missed by genome-wide association studies. Furthermore, many SNPs without a causal effect may show a strong association with a phenotype and, therefore, enter as surrogates in the model above for genotypes at causal loci.

A good polygenic score is highly correlated with the phenotype in question. The main factors affecting this prediction correlation are the trait heritability, the size of the dataset used to train the polygenic score, and the number of SNPs available for the analysis. Using extensive training datasets increases the power to detect associations (*see the example on page 289*) and the accuracy of SNP effect estimates, thus improving the prediction correlation between polygenic scores and phenotypes.

tion, while having intermediate frequencies in another population. Such alleles contribute to phenotypic variance in the population where such loci are segregating, but will not do so in a population in which only one allele is present. Furthermore, although Fisher's additive model does not account for gene–gene interactions, it is well known that such interactions exist. For example, one gene can mask, promote, or suppress the expression of another gene—a form of *epistasis*, a term that refers to gene–gene interactions. Under epistasis, the additive effects of alleles (the α's in the Fisher–Wright equation) become background-dependent. In other words, they can be expected to be different in two populations if the allele frequencies at interacting loci vary between populations. Likewise, gene-by-environment interactions are also ignored in Fisher's model and may make genetic effects on a trait population-specific if environmental factors (such as diet, exercise, climate, or access to health care) vary between populations.

Finally, differences in linkage disequilibrium patterns between populations may also lead to heterogeneity in SNP effects (the β's in the linear equation [*see facing page*] used in genome-wide association studies) over populations, even when the genetic causal structure in terms of effects and frequencies is the same in both populations.

How much of the low portability of polygenic scores between populations is due to heterogeneity at the causal level and how much is due to imperfect linkage disequilibrium with markers remains to be determined. It is clearly a difficult and crucial problem. However, recent research suggests that imperfect linkage disequilibrium may be a major factor affecting the poor portability of polygenic scores between populations.

Machine Learning in Genomics

The Fisher–Wright model has provided a powerful tool to describe and predict patterns of phenotypic variation from genetic information in the form of a pedigree or DNA sequences. This usefulness does not imply that the model is a good tool for learning about function, because the Fisher–Wright model is based on assumptions that have been falsified by many known facts about biochemistry and molecular biology, including gene-by-gene and gene-by-environment interactions.

Considering the success of machine learning methods in many fields, one may wonder whether some machine learning methods such as deep neural networks (also called deep learning) will eventually replace the Fisher–Wright model for genomic prediction. In principle, deep learning models can capture complex nonlinear patterns (for example, gene-by-gene interactions or sex and ancestry differences in the DNA-phenotype map) that a linear model may miss. But the empirical evidence available, especially from animal and plant breeding, suggests that machine learning methods do not systematically outperform additive models for genomic prediction—and often underperform. Why this underperformance happens is not entirely clear. Some may argue that larger sample sizes may be needed to better calibrate deep learning models. However, even studies using a large sample size have not shown a clear superiority over the additive model in terms of predictive ability.

The success of deep learning stems in part from its ability to "engineer features"—that is, to extract features from inputs (such as SNPs) that are correlated with an outcome. This method should benefit genomic prediction because, for example, a good deep learning model may be capable of identifying sets of interacting genes and using them to predict phenotypes. However, in quantitative genomics the input space (number of SNPs) and the number of possible DNA-phenotype maps that one can build are too large to be explored effectively by current algorithms.

Machine learning models can capture nonlinear patterns for predictors with sizable effects. As noted earlier, individual SNPs have limited explanatory power, which may explain why machine learning has yet to outperform linear models. However, recent studies suggest that these methods may be more helpful in developing models that integrate SNPs with other predictors that have sizable effects and may modulate SNP effects such as sex, age, and ancestry.

Given the success of machine learning in other fields, the continued improvement of algorithms and software, and the increasing availability of large genomic datasets, the number of studies exploring machine learning for genomic prediction will surely increase. They may produce knowledge that leads to more intensive use of machine learning for predicting phenotypes and disease risk using DNA. Until that happens, a century-old model remains our best genomic prediction tool.

Bibliography

Bellot, P., G. de los Campos, and M. Pérez-Enciso. 2018. Can deep learning improve genomic prediction of complex human traits? *Genetics* 2010:809–819.

de los Campos, G., D. A. Sorensen, and D. Gianola. 2015. Genomic heritability: What is it? *PLoS Genetics* 11:e1005048.

de los Campos, G., D. A. Sorensen, and M. A. Toro. 2019. Imperfect linkage disequilibrium generates phantom epistasis (and perils of big data). *G3 Genes|Genomes|Genetics* 9:1429–1436.

Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* 52:399–433.

Hill, W. G., M. E. Goddard, and P. M. Visscher. 2008. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics* 4:e1000008.

Lello, L., S. G. Avery, L. Tellier, A. I. Vazquez, G. de los Campos, and S. D. H. Hsu. 2018. Accurate genomic prediction of human height. *Genetics* 2:477–497.

Maher, B. 2008. Personal genomes: The case of the missing heritability. *Nature* 456:18–21.

Meuwissen, T. H., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.

Privé, F., et al. 2022. Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. *American Journal of Human Genetics* 109:12–23.

Sirugo, G., S. M. Williams, and S. A. Tishkoff. 2019. The missing diversity in human genetic studies. *Cell* 177:26–31.

Wang, W. Y. S., B. J. Barratt, D. G. Clayton, and J. A. Todd. 2005. Genome-wide association studies: Theoretical and practical concerns. *Nature Reviews Genetics* 6:109–118.

Wright, S. 1921. Systems of mating. *Genetics* 6:111–178.

Gustavo de los Campos is a professor of epidemiology & biostatistics and statistics & probability, and a member of the Institute of Quantitative Health Sciences and Engineering at Michigan State University. His research centers on developing and applying statistical models and software for the genetic analysis and prediction of complex traits using DNA. Daniel Gianola is an emeritus professor of animal breeding and quantitative genetics and biostatistics at the University of Wisconsin, Madison. His research centers on the theory of quantitative genetics and on statistical genetics. He has been one of the main contributors to the adoption of Bayesian methods for analyzing and predicting complex traits in quantitative genetics. Email for de los Campos: gustavoc@msu.edu

Medical Imaging in Increasing Dimensions

Combining technologies and moving into virtual space makes seeing into the body more accurate and beneficial.

Ge Wang

In May 2022, I had a dizzy spell and went to Albany Medical Center. Worried that I might be having a stroke, my care team ordered computed tomography (CT) and magnetic resonance imaging (MRI) scans. Both are needed to determine whether a patient should receive thrombolytic therapy to destroy blood clots (if brain vessel blockage is shown by CT) or other interventions to save neurologic functions (as evaluated by MRI). I had the CT scan first, which took only a few seconds, but I had to wait until the next day for the MRI scan, which took more than 20 minutes.

Fortunately, both scans showed healthy brain vasculature and tissues. I was lucky that my case wasn't a real emergency, or my results might have come too late to be useful. As a researcher who has been working on medical imaging for decades, I was especially reassured by seeing my CT and MRI results. These scans are the eyes of modern medicine, noninvasively peeking into the human body and rapidly making a patient virtually transparent to reveal subtle features in a region of interest.

Current imaging exams are built on a deep foundation of development to make them the reliable medical workhorses we know them to be today. But as my personal experience demonstrated, there's still a lot of room for developing better and faster imaging techniques to help patients. Over the past three decades, my colleagues and I have been working to make those advances happen, by adding new dimensions to the ways we view the body.

2D Imaging with X-rays

Modern CT scans are the culmination of a series of imaging technologies that began with the 1895 discovery of x-rays by Prussian physicist Wilhelm Röntgen. In 1901, he won a Nobel Prize for this discovery, which soon revolutionized medical diagnoses.

Although visible light cannot penetrate our bodies, an x-ray beam can pass through us easily. During the x-ray imaging process, an x-ray beam's intensity is *attenuated*, or dampened, by the material elements it meets along its path. The degree of attenuation depends on the type of material, such as bones (more attenuating) or soft tissues (less attenuating). An x-ray image is produced from many x-rays that emanate from a single source and illuminate an x-ray sensitive film, screen, or detector array. Much like the way that pixels together compose an image on a cell phone screen, the collection of detected x-ray attenuation signatures combine to form a picture of the body, rendering bright, white skeletons (low exposure) against dark, murky masses of tissue (higher exposure).

In the late 1960s, Godfrey Hounsfield, an engineer working on radar and guided missiles at the British conglomerate EMI, and Allan Cormack, a Tufts University mathematician, developed a way to put together multiple one-dimensional x-ray profiles from various angles to generate a cross-sectional image inside a patient. They called it computerized axial tomography (CAT), also referred to as computed tomography (CT). Tomography derives from the Greek word *tomos*, or section. A traditional CT scan combines many x-ray profiles by moving an x-ray source and detector assembly around a patient's body, then computationally reconstructing two-dimensional images of the internal organs. Hounsfield and Cormack won a Nobel Prize in 1979 for their work.

Two-dimensional CT images offered a tremendous improvement on traditional x-ray radiograms where anatomical structures are overlapped along x-ray paths, but within decades the capabilities of medical imaging had been dramatically increased again by adding additional image dimensions.

In the nearly 130 years since the first x-ray image was taken, researchers have developed a host of medical imaging options, each suited to seeing specific dimensions of the body. By combining multiple technologies, clinicians could one day access complete, real-time information about a living human body. In order of their first widespread clinical appearances are x-rays (1890s), computed tomography (CT, 1970s), magnetic resonance imaging (MRI, 1980s), and positron emission tomography (PET, 1990s).

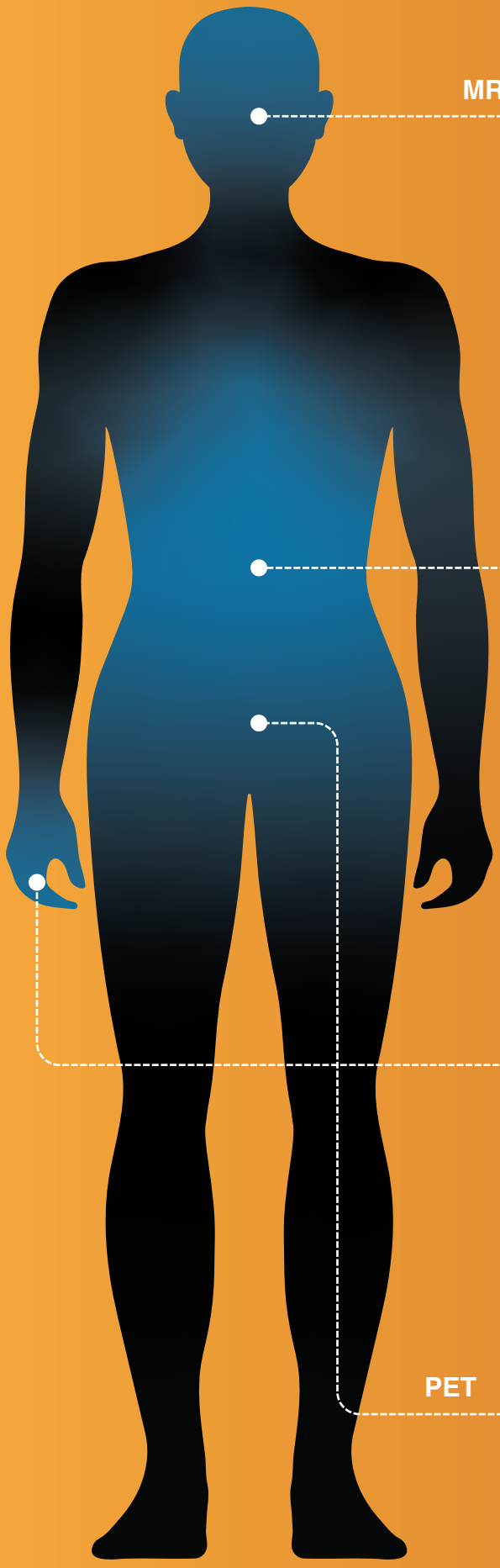
QUICK TAKE

Medical imaging has come a long way since the first x-ray photo in 1895, and can now produce dynamic 3D models of patients' bodies to aid doctors' diagnoses.

The more types of technologies combined in medical scans, the more dimensional views and scales of details become available in the resulting images.

The next step is to combine all relevant imaging modalities into a single machine that can view patients from all angles, perhaps even directing AI-enhanced remote surgeons.

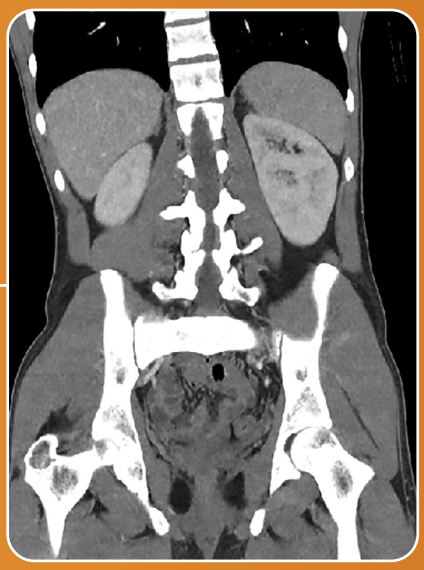
Barbara Aulicino; medical scans from Wikimedia Commons: deradrian; Mikael Häggström, M.D.; Photojack50; Soichi Oka, Kenji Ono, Kenta Kajiyam, and Katsuma Yoshimatsu



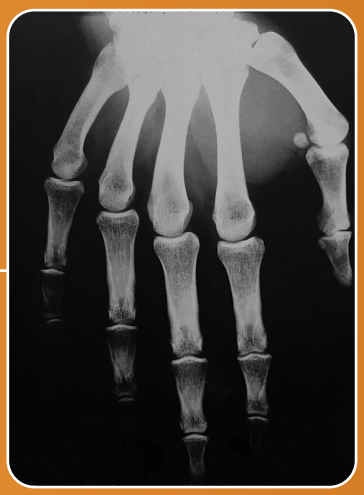
MRI



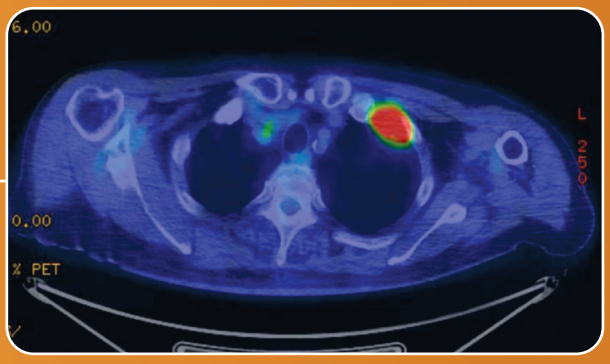
CT



x-ray

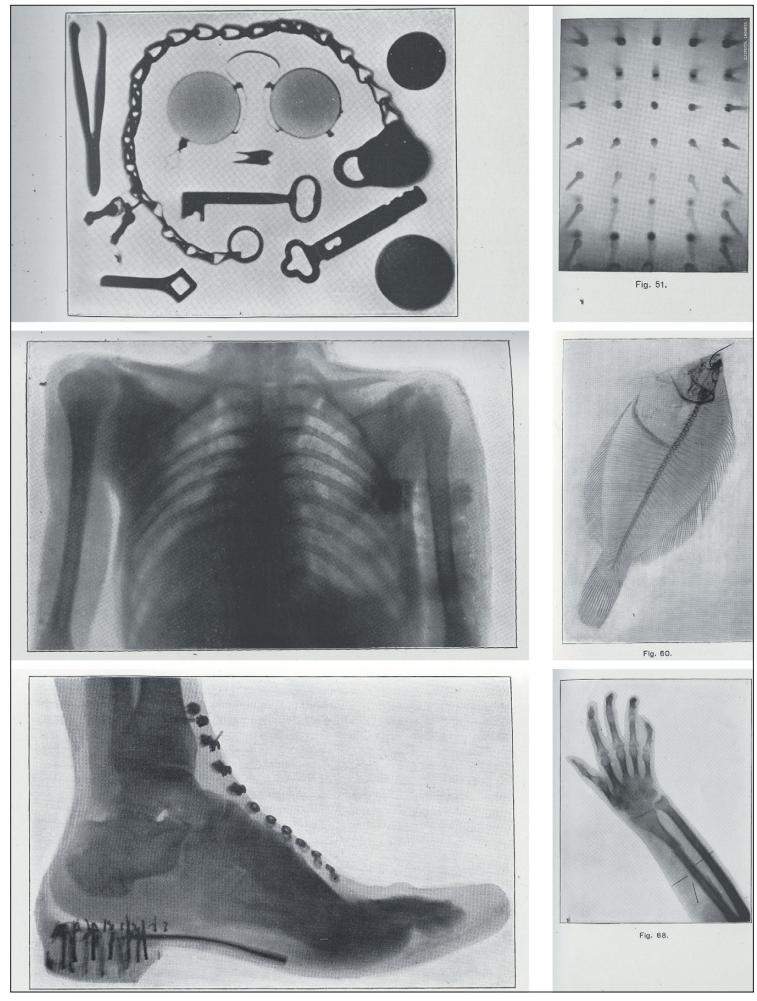


PET





Wikimedia Commons/Wellcome Images



National Library of Medicine/History of Medicine

"Hand mit Ringen,"or "Hand with Ring," by Wilhelm Röntgen, 1895 (*above*), was the first x-ray image ever produced. It features the bone of his wife's left hand with her wedding ring. Curious researchers then tested the technology on myriad items, including tools (*top middle*), nails (*top right*), fish (*center right*), and various human body parts, such as arms (*bottom right*), feet (*bottom middle, inside a boot*), and chests (*center*).

Each new dimension represents a breakthrough in our ability to perform medical diagnoses and treat patients.

From 2D to 4D Imaging

Both the first x-ray radiogram and the first CT images were two-dimensional, which limited their ability to capture our three-dimensional anatomy. Modern CT scanners generate 3D images by working in what's called a spiral or helical *cone-beam scanning mode*. In that mode, an x-ray source and a 2D detector array rotate together while the patient is moved through the CT scanner gantry. This process results in a spiral trajectory of the x-ray focal spot relative to the patient, which produces a cone-shaped beam. Then the collected cone-shaped x-ray beam data are used to reconstruct a volumetric (3D) image within seconds.

These scans are the eyes of modern medicine, noninvasively peeking into the human body and rapidly making a patient virtually transparent to reveal subtle features in a region of interest.

In the early 1990s, my colleagues and I formulated the first spiral cone-beam CT image reconstruction algorithm. Along with many of our peers, we later developed a number of more refined algorithms for this purpose. Although our first spiral cone-beam algorithm performed only approximate image reconstruction, modern algorithms enable highly accurate image reconstruction of much-improved image quality at much-reduced radiation dosages. A scan today can generate a 3D model of a patient's heart that includes color-coded identification of different plaques and tissues along with detailed information about blood flow and blockages in arteries; intra-abdominal scans can detect abnormalities with 95 percent accuracy. In total, there are about 200 million CT scans worldwide annually, most of which use the spiral cone-beam mode.

Efforts to develop better treatments for cardiac disease, the leading cause of death worldwide, prompted the push for four-dimensional CT scans—that is, 3D scans that incorporate the added dimension of time. Solving the problem

of capturing rapid and irregular heart motion remained elusive until the early 2000s with the development of accelerated CT scanning and advanced image reconstruction algorithms aided by an individual electrocardiogram (ECG), which records heart activity in a graph. Both technologies were major engineering challenges. In order to capture the rapid motion of a beating heart, scanning needed to happen much faster than was possible before.



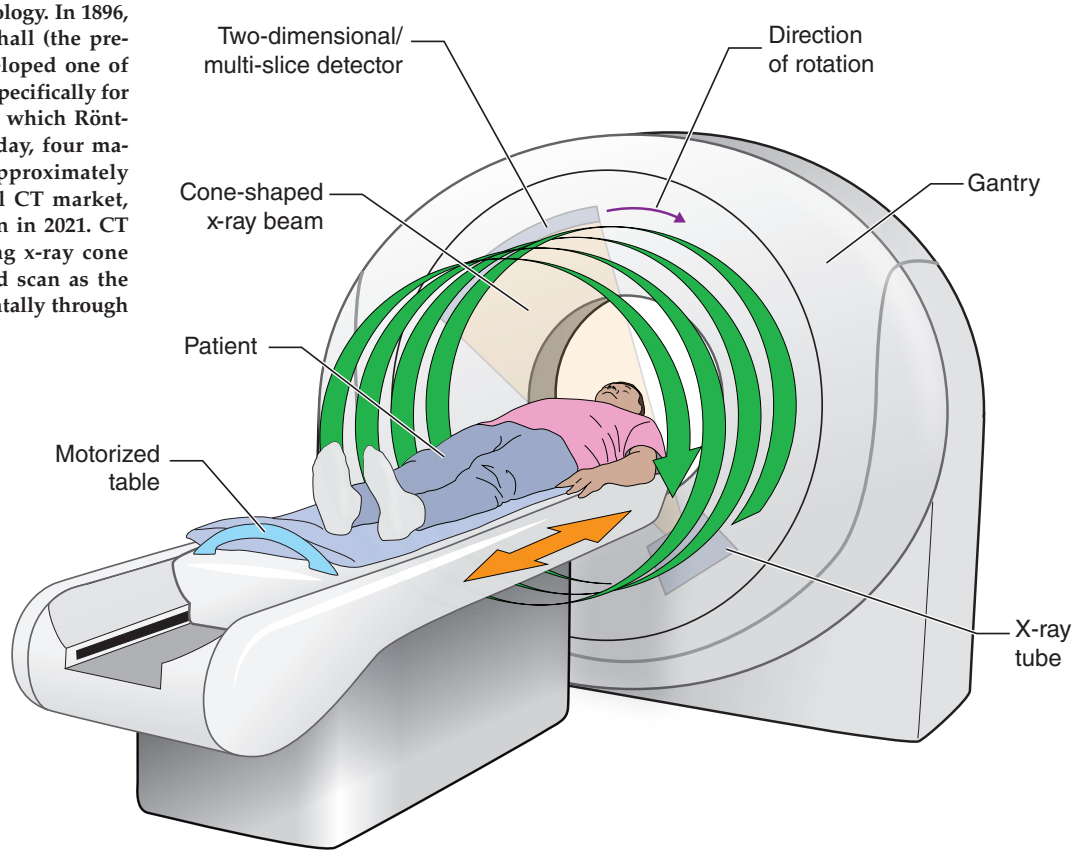
Wikimedia Commons

But spinning the CT gantries at higher rates resulted in extreme g-forces that required new, high-precision hardware. Then engineers needed to create more sophisticated algorithms to solve the mathematical problem of how to accurately match data segments with cardiac cycles.

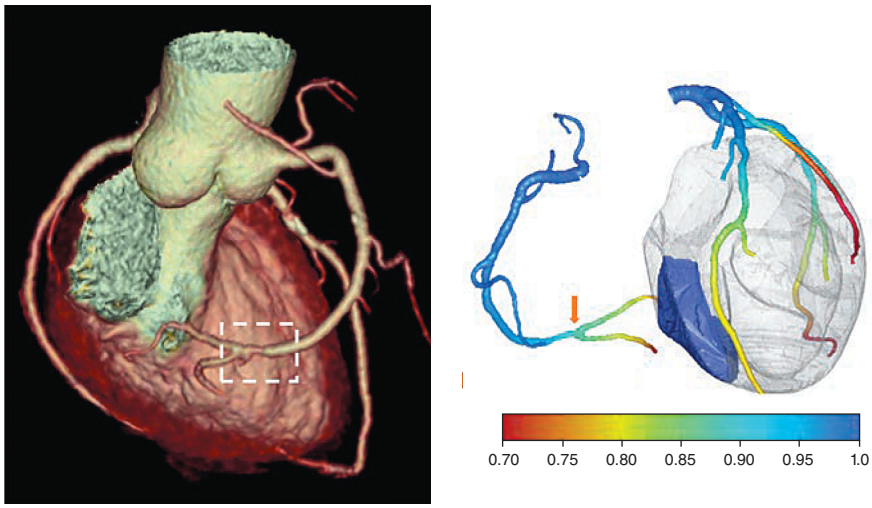
With the new hardware, the x-ray source and detector could rotate three to four turns per second while the ECG-aided algorithms synchronized the collected data segments to their corresponding heartbeat phases in order to reconstruct an image at each ECG phase. As a result, human cardiac motion could in many cases be effectively frozen in time, revealing such features as cardiac vessel narrowing and calcification.

For example, from a coronary CT scan, an algorithm can build an individualized blood flow model of the beating heart. It does this by solving partial differential equations, which involve rate changes that depend on multiple variables, in order to describe blood flow through biological tissues and to extract physiological parameters of the

Modern CT scanners, such as this Siemens machine (*above*), build off of 19th-century x-ray technology. In 1896, Reiniger, Gebbert & Schall (the precursor to Siemens) developed one of the first tubes designed specifically for producing x-ray images, which Röntgen later purchased. Today, four major players command approximately 70 percent of the global CT market, estimated at $6.39 billion in 2021. CT machines use a spinning x-ray cone to create a spiral-shaped scan as the patient is moved horizontally through the machine (*right*).



Jason McAlexander



Dr. Ofer Benjaminov, Rabin Medical Center, Israel

Combining information from a CT scan (*left*) and a patient's heartbeats (from an electrocardiogram) allows health care providers to view personalized, 3D models of the patient's arteries and any blockages (*right*). The colors in the model represent *fractional flow rates*, indicating degrees of blockage; red represents lower flow, meaning that section is more blocked than the blue sections.

coronary arteries. Because every patient is different, a personalized model can help optimize diagnosis and tailor treatment for each individual. Cardiac CT has now become the first-line imaging technique for patients with suspected coronary artery diseases.

A prerequisite for precision medicine is to have an understanding of where, and for how long, all of the roughly 100,000 types of key proteins are in the human body.

The Fifth Dimension

Going beyond 4D, biomedical engineers have added a fifth dimension to CT by incorporating spectral data, considering the wavelengths of the x-ray radiation used to create images. This so-called *photon-counting CT* creates images that are precisely colorized to differentiate between biological materials as well as to track injected contrast agents.

Previously, all x-ray radiograms and CT images were in grayscale. But the potential for color was there. Just as an incandescent light emits a whole range of colors, from longer red wavelengths to shorter purple ones, a common x-ray tube emits photons over a wide spectrum of x-ray "colors," with wavelengths in a range of nanometer lengths. The most common type of CT scanner, however, uses an energy-integrating detector array that accumulates all the energy deposits from incoming x-ray photons together into a single total. As a result, the differences between the energy-specific deposits are lost in translation at the final image, leaving just shades of gray.

Over the past several years, a major technological frontier in the CT field has been the photon-counting CT detector, which individually records x-ray photons along with their wavelengths or energy levels, from soft x-rays (longer wavelengths and less energetic) to hard x-rays (shorter wavelengths and more energetic). In this way, the photon-counting detector can sense the whole x-ray spectrum. The transition from grayscale CT to photon-counting spectral CT is happening now in much the same way as the transition from grayscale to color television happened almost a century ago.

Five-dimensional scanners use two main types of photon-counting detectors: cadmium zinc telluride–based ones used by MARS Bioimaging and Siemens, and silicon-based ones used by General Electric. The former type is much more compact because of its higher-density sensing material, which enables it to absorb and detect x-rays with thinner sensors. However, the latter type is more reliable because it is based on more mature technology that has been used for decades in the semiconductor industry.

The major clinical advantage of either type of photon-counting spectral CT is its ability to distinguish between similar but subtly different materials. For example, it can better differentiate stable plaques from risky ones in main blood vessels that are responsible for heart attacks and strokes. High-risk or vulnerable plaque has three signs: a lipid core (submillimeter to a few millimeters in size), a thin fibrous cap (less than 100 micrometers in thickness), and microcalcifications (on the order of tens of microns). Stable plaques do not contain these structures. X-rays absorb in each of these materials differently depending on their characteristics and the photon energies used in the scan.

Molecular and Cellular Imaging

A prerequisite for precision medicine is to have an understanding of where, and for how long, all the roughly 100,000 types of key proteins are in the human body. Each protein adds another type of information, and hence another dimension, to the scan. To further improve CT imaging, an increasing number of contrast agents and nanoparticles are being developed to be injectable into humans and animals for more accurate and reliable quantification of biologically relevant features, including proteins. These agents, like bones and tissues, each have their own spectrally distinct attenuation signatures, and they can be selectively bound to molecular and cellular features or accumulated in vasculature to enhance image-based biomarkers.

For instance, a cancer patient undergoing therapy today can be injected with various nanoparticles to label cancer cells and therapeutic drugs. Then a nanoparticle-enhanced CT scan can reveal their distributions in the patient's body. A major advantage of this method over previous anatomical-only imaging is that it can more accurately distinguish malignant from benign tumors. This method differs

from photon-counting CT scans because benign and malignant tumors might have similar densities and thus appear to be similar based on photon counting, whereas a nanoparticle can be tagged to only cancerous cells.

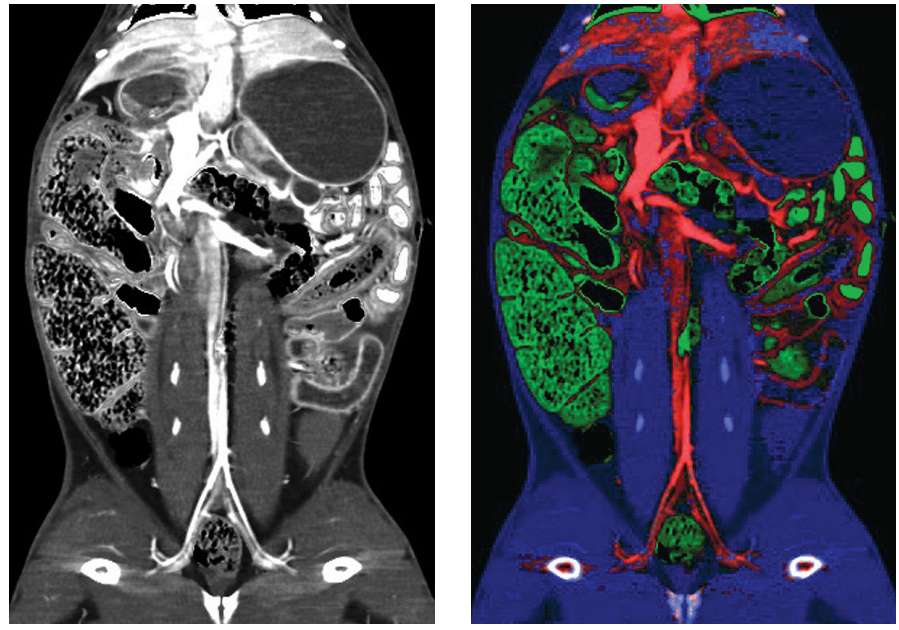
Combining Technologies

CT alone is insufficient to image all functions and forms of biological targets in a patient. Inspired by x-ray radiography and CT, other researchers have developed a number of additional imaging technologies that are now in widespread use, including magnetic resonance imaging (MRI), positron emission tomography (PET), single-photon emission computed tomography (SPECT), ultrasound imaging, optical coherence tomography (OCT), and photo-acoustic tomography (PAT). These various types of scans view the body through different lenses in terms of their unique physical interactions with biological tissues.

In the medical imaging field, researchers have been combining two or more of these technologies to perform what's known as *multimodality imaging*. The first medical multimodality imaging success was the PET-CT scanner in 1998. In this type of machine, functional information from PET is labeled with a radioactive tracer in cancer regions that can be overlaid onto the anatomical structure defined by CT. More recently, Siemens designed the first PET-MRI scanner for the same purpose, because MRI images contain much richer soft tissue information than CT. This advance has enabled or improved important clinical applications, such as eliminating radiation doses from CT and allowing doctors to identify the stage of progression for ailments including cancer, neurological disorders, and heart disease, all with greater accuracy than traditional imaging techniques.

As a long-term and rather challenging goal, my colleagues and I hope to develop *omni-tomography*, for all-in-one and all-at-once imaging. All-in-one means that all tomographic imaging modalities are integrated into a single machine gantry. All-at-once means that diverse datasets can be collected simultaneously. This hybrid data stream would automatically register in space, time, and diverse contrasts to provide even higher-dimensional—and therefore more informative—imaging.

True omni-tomography is not yet possible, but as a step in that direction my collaborators and I have been promoting the idea of simultaneous CT-MRI. Such a CT-MRI machine would be ideal for an acute stroke patient, because the window of opportunity for effective treatment is within a few hours. In addition to speeding up diagnoses, simultaneous CT-MRI would combine dimensions of both images, reduce misregistration (image misalignment) between them, and optimize diagnostic performance.



Courtesy of Benjamin Yeh, MD, UCSF

Spectral CT creates images that are precisely colorized to differentiate between biological materials as well as to track injected biomarkers, as demonstrated by these two versions of scans of a rabbit's midsection. In grayscale, the contrast agents in different regions are indistinguishable, whereas in color the silicon contrast displays as green, the iodine contrast as red, and the soft tissues as blue.

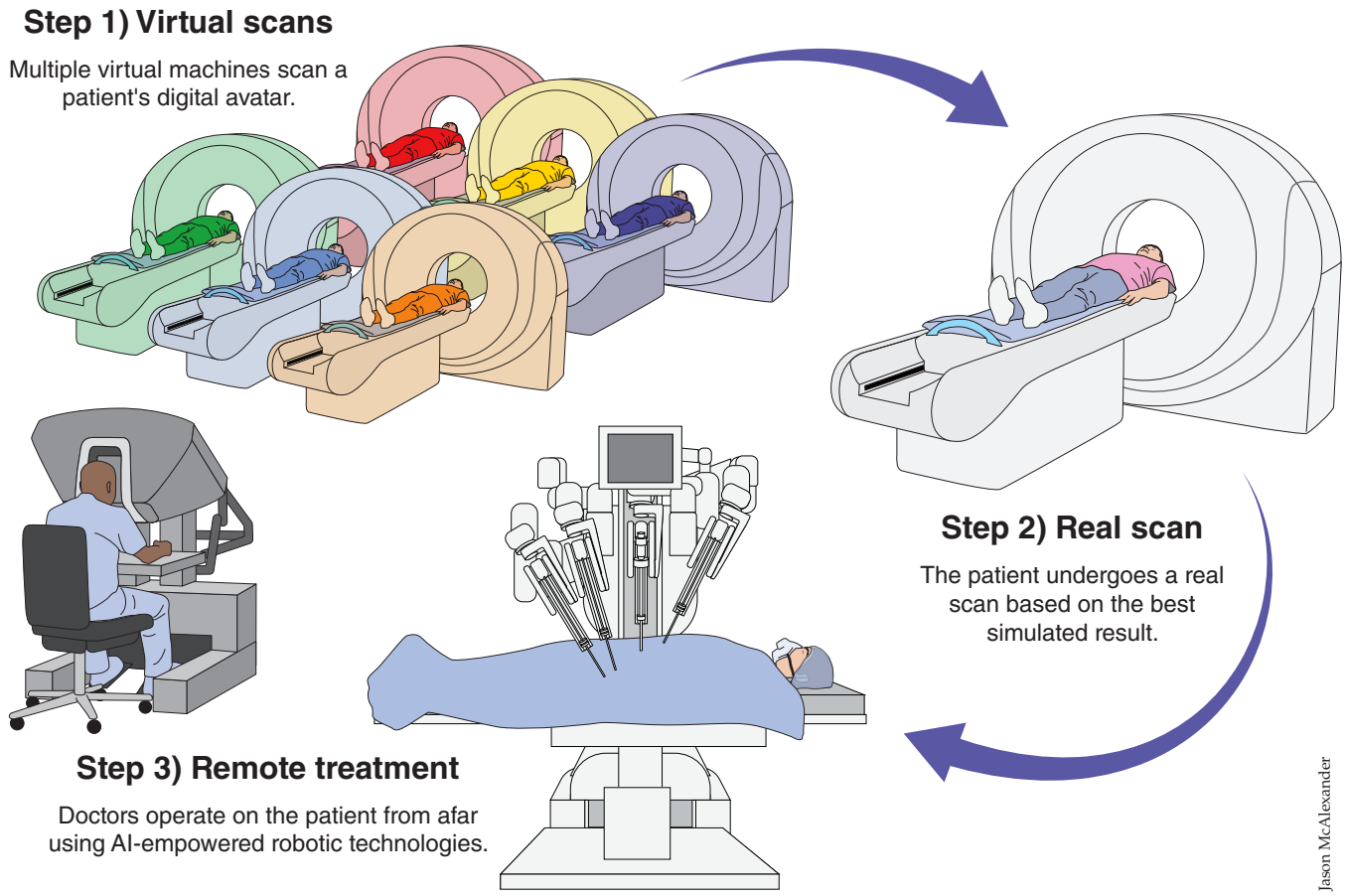
Adding imaginary dimensions is common in mathematics, but it would be revolutionary to medical imaging.

For truly comprehensive biomedicine analysis, we need to perform imaging studies at various scales. For example, x-ray imaging can cover six orders of magnitude in terms of image resolution and object size, from nanometer-sized subcellular features to cells and meter-sized human bodies. The case is similar with MRI, ultrasound, and other imaging modalities. This scale dimension is critically important—for example, to link a genetic change to a physiological outcome—because biological architectures are intrinsically organized at multiple scales.

Artificial intelligence software that can determine the 3D structures of proteins from their amino acid sequences is now widely available—notably, Google DeepMind made its AlphaFold2 system free and open source in 2021. The advent of such powerful new tools suggests the enormous potential of medical imaging to produce much higher dimensional and much more complex datasets for us to look into life's innermost secrets.

Virtual Dimensions

Now that medical imaging can provide multiple dimensions of information on the patient, the three spatial dimensions and one temporal dimension we perceive might seem to be lim-



The imagined healthcare metaverse of the author and his colleagues is referred to as *MeTAI* (Medical Technology and Artificial Intelligence). Among other data integration and computational advantages, a health care metaverse would enable virtual scanning of a patient with all available technologies before subjecting them to an in-person, or "real," scan. This approach would reduce overhead costs, time to treatment, and radiation doses.

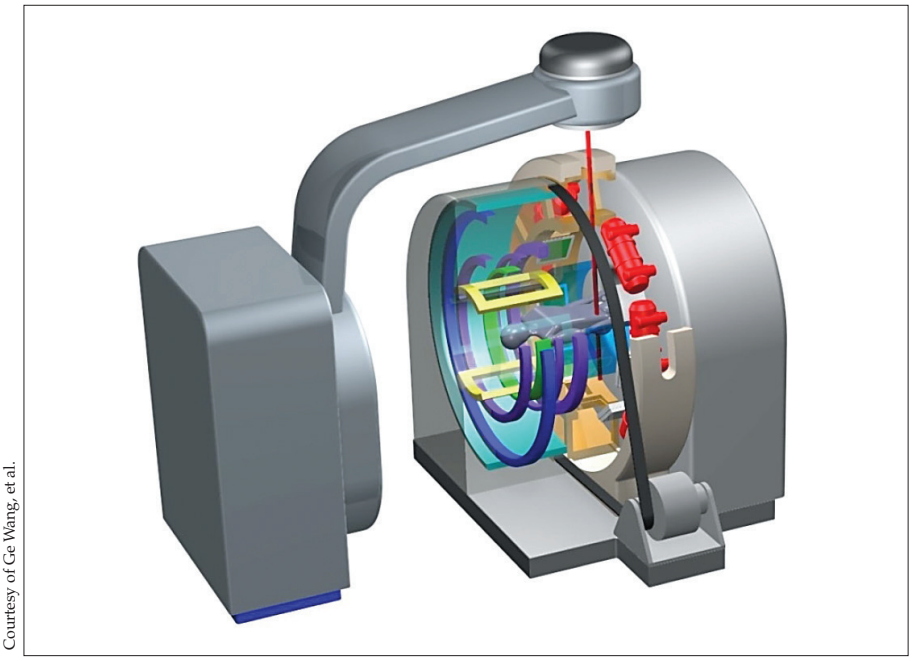
iting factors in interpreting all those data. One way to compensate for that is to combine medical scans from the real world with virtual datasets to create an augmented view of the human body. Doing so relies on using the imaginary number space in addition to the typical x-, y-, and z-axes we use to describe the "real" space of our day-to-day lives. This technique draws on the same types of technologies that define, among other things, virtual worlds like those we inhabit in online gaming.

Looking forward, a number of cutting-edge technologies could come together to make medical imaging faster, more informative, and ultimately more lifesaving.

Information from virtual worlds, known in mathematics as *complex-valued space*, could enhance medical imaging through the so-called *metaverse*. In the metaverse, we could make digital twins of existing physical scanners and avatars of patients. Before the patient undergoes a real physical scan, their avatar could be modified to have different pathological features of relevance and scanned on a digital version of each scanner to determine which imaging technique would be most useful.

For example, a patient suspected of having a coronary artery disease might have their avatar modified to display a pathology their physician would expect to find, such as narrowed vessels. After putting the avatar through a virtual scan on every type of machine, the corresponding AI software reconstructs and analyzes the images. Once the patient and their care team determine which result is most informative, an actual physical scan can take place. Finally, if it is clinically indicated, the patient could be automatically and even remotely operated on in a robotic surgical suite or radiotherapy facility.

In addition to making diagnosis and treatment more personalized, effective, and efficient, with appropriate security and permission, the images, tomographic raw data, and other relevant information could be made available



Courtesy of Ge Wang, et al.

This schematic, created by the author and his colleagues, depicts their proposed combined CT-MRI machine. CT scans offer higher levels of detail than MRI and can spot issues such as blood clots, subtle fractures, and organ injuries; MRI scans are better at spotting soft tissue problems, such as sports injuries, or neural degenerative conditions. Performing these scans together would speed up time to treatment, a particular concern for acute emergencies such as strokes.

to researchers and regulatory agencies. Then another revolution would be possible: All these simulated and real images could be used in augmented clinical trials, perhaps to enhance the statistical power of the AI machines or to match similar cases to aid in health care decision-making.

The future health care metaverse that we envision would be a complex-valued mathematical space, orders of magnitude more powerful and informative than our real, 3D, physical space. Looking forward, I see a number of cutting-edge technologies and unprecedented ideas from many disciplines coming together to make medical imaging faster, more informative, and ultimately more lifesaving. A health care metaverse awaits where the boundaries between real and virtual worlds blur, and where AI models will enable intelligent medicine and prevention wherever and whenever it's needed, with maximum accessibility for all patients.

References

Jaffe, C. 1982. Medical imaging. *American Scientist* 70:576–585.

Wang, G. 2015. How x-rays see through your skin. TED-Ed. YouTube (posted June 22). youtu.be/gsV7SJDDCY4

Wang, G., et al. 2022. Development of metaverse for intelligent healthcare. *Nature Machine Intelligence* 4:922–929.

Ge Wang is Clark-Crossan Chair Professor and director of the Biomedical Imaging Center at Rensselaer Polytechnic Institute. He pioneered the cone-beam spiral CT method in the early 1990s. Email: wangg6@rpi.edu

Curious about becoming a member of Sigma Xi?

Since the Society's founding in 1886, induction into Sigma Xi has been a career catalyst and milestone for hundreds of thousands of distinguished scientists and engineers.

Visit sigmaxi.org to nominate yourself or others you know who hold the key to research excellence!

SIGMA XI
THE SCIENTIFIC RESEARCH HONOR SOCIETY, INC.



Time Is an Object

Not a backdrop, an illusion, or an emergent phenomenon, time has a physical size that can be measured in the laboratory.

Sara Walker and Lee Cronin

A timeless universe is hard to imagine, but not because time is a technically complex or philosophically elusive concept. There is a practical reason: The act of imagining timelessness requires time to pass. Even when you try to imagine its absence, you sense it moving as your thoughts shift, your heart pumps blood to your brain, and images, sounds, and smells move around you. The thing that is time never seems to stop. You may even feel woven into its ever-moving fabric as you experience the universe coming together and apart. But is that how time really works?

According to Albert Einstein, our experience of the past, present, and future is nothing more than "a stubbornly persistent illusion." According to Isaac Newton, time is nothing more than a backdrop, outside of life. And according to the laws of thermodynamics, time is nothing more than entropy and heat. In the history of modern physics, there has never been a widely accepted theory in which a moving, directional sense of time is fundamental. Many of our most basic descriptions of nature—from the laws of movement to the properties of molecules and matter—seem to exist in a universe where time doesn't really pass. However, recent research across a variety of fields suggests that the concept of time might be more important than most physicists once assumed.

A new form of physics called *assembly theory* suggests that a moving, directional sense of time is real and fundamental. It suggests that the complex objects in our universe that have been made by life, including microbes, computers, and cities, do not exist outside of time: They are impossible without the movement of time. From this perspective, the passing of time is not only intrinsic to the evolution of life or our experience of the universe. It is also the ever-moving material fabric of the universe itself. Time is an object. It has a physical size, like space. And it can be measured at a molecular level in laboratories.

The unification of time and space radically changed the trajectory of physics in the 20th century. It opened new possibilities for how we think about reality. What could the unification of time and matter do in our century? What happens when time is an object?

Incompatible Ideas of Time

For Newton, time was fixed. In his laws of motion and gravity, which describe how objects change their position in space, time is an absolute backdrop. Newtonian time passes, but never changes. And it's a view of time that endures in modern physics—even in the wave functions of quantum mechanics time is a backdrop, not a fundamental feature. For Einstein, however, time was not absolute. It was relative to each observer. Einsteinian time is what is measured by the ticking of clocks; space is measured by the ticks on rulers that record distances. By studying the relative motions of ticking clocks and ticks on rulers, Einstein was able to combine the concepts of how we measure both space and time into a unified structure we now call *space-time*. In this structure, space is infinite and all points exist at once. But time, as Einstein described it, also has this property, which means that all times—past, present, and future—are equally real. The result is sometimes called a *block universe*, which contains everything that has and will happen in space and time. Today, most physicists support the notion of the block universe.

But the block universe was cracked before it even arrived. In the early 1800s, nearly a century before Einstein developed the concept of space-time, French engineer Nicolas Léonard Sadi Carnot and other physicists were already questioning the notion that time was either a backdrop or an illusion. These questions persisted as 19th century scholars such as Austrian physicist Ludwig Boltzmann also began to turn their minds to the problems that came with a new kind of technology: the engine.

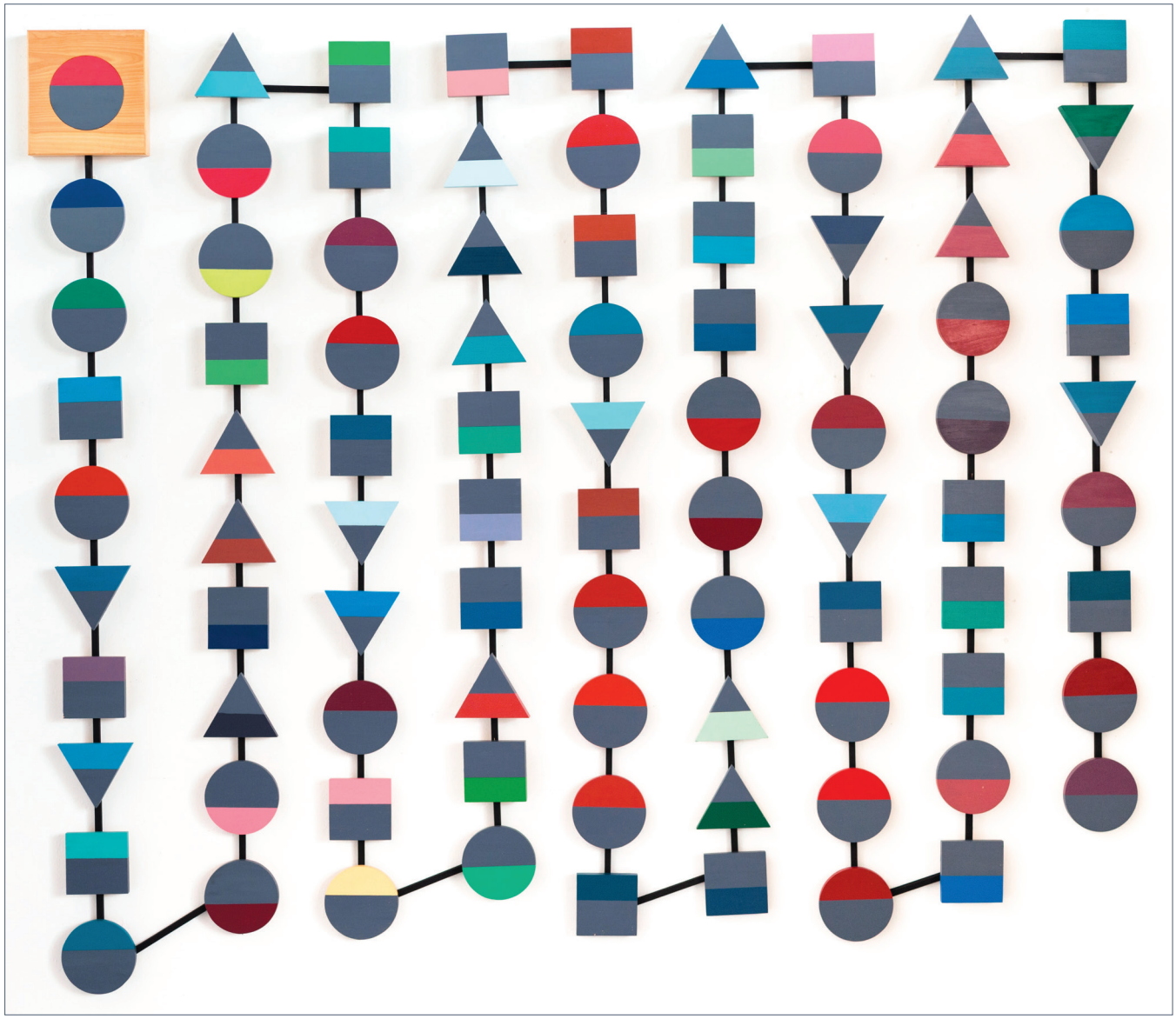
Though engines could be mechanically reproduced, physicists didn't know exactly how they functioned. Newtonian mechanics were reversible; engines were not. Newton's Solar System ran equally well moving forward or backward in time. However, if you drove a car and it ran out of fuel, you could not run the engine in reverse, take back the heat that was generated, and unburn the fuel. Physicists in that era suspected that engines must be adhering to certain laws, even if those laws were unknown. What they found was that engines do not function unless time passes and has a direction. By exploiting differences in temperature, engines drive the movement of heat from warm parts to cold parts. As time moves forward, the temper-

QUICK TAKE

Scientists have no consistent definition of time. In some fields of study, it is regarded as a fundamental aspect of nature; in others, as an illusion of human experience.

***Assembly theory* introduces** a unified conception of time by linking it to the influence of life: The flow of time is defined by the way that complex systems come into being.

By quantifying complexity, the *assembly index* identifies systems that are uniquely associated with life. This method could assist the search for life beyond Earth.



Carol Ladewig; photo: Dana Davis Photography

Carol Ladewig's art installation series "Painting Time" (including "Winter 2016," *above*) depicts time the way humans experience it: as a linear sequence of events. But that experience runs contrary to the fundamental laws of physics. *Assembly theory* attempts to reconcile the biological and physical conceptions of time—and also how we define life itself.

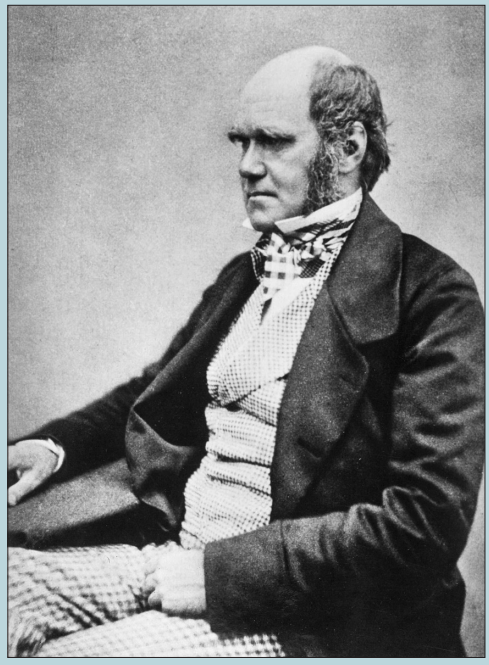
ature difference diminishes and less "work" can be done. This principle is the essence of the second law of thermodynamics (also known as the law of entropy), which was proposed by Carnot and later explained statistically by Boltzmann. The law describes the way that less useful "work" can be done by an engine over time. You must occasionally refuel your car, and entropy must always increase.

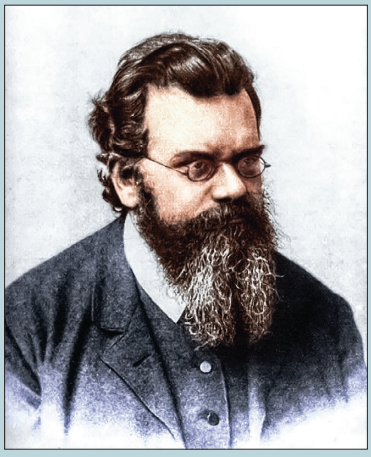
This makes sense in the context of engines or other complex objects, but it is not helpful when dealing with a single particle. It is meaningless to talk about the temperature of a single particle, because temperature is a way of quantifying the average kinetic energy of many particles. In the laws of thermodynamics, the flow and directionality of time are considered an emergent property rather than a backdrop or an illusion—a property associated with the behavior of large numbers of objects. Although thermodynamic theory introduced how time should have a directionality to its passage, this property was not fundamental. In physics, "fundamental" properties are reserved for those properties that cannot be described in other terms. The arrow of time in thermodynamics is therefore considered "emergent" because it can be explained in terms of more fundamental concepts, such as entropy and heat.

Charles Darwin, working between the steam engine era of Carnot and the emergence of Einstein's block universe, was among the first to clearly see how life must exist in time. In the final sentence from *On the Origin of Species* (1859), he eloquently captured this perspective: "Whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been and are being evolved." The arrival of Darwin's "endless forms" can be explained only in a universe where time exists and has a clear directionality.

During the past several billion years, life has evolved from single-celled organisms to complex multicellular organisms. It has evolved from simple societies to teeming cities, and now a planet potentially capable of reproducing its life on other worlds. These things take time to come into existence, because they can emerge only through the processes of selection and evolution.







Wikimedia Commons

In the 19th century, physicists such as Nicolas Léonard Sadi Carnot (*top left*) and Ludwig Boltzmann (*bottom left*) developed theories to explain why some processes cannot be reversed, even though the fundamental laws of physics do not seem to depend on the direction of time. Meanwhile, Charles Darwin (*above*) was contemplating the theory of evolution, in which the emergence of "endless forms" of new types of life depends on the forward passage of time.

We think Darwin's insight does not go deep enough. Evolution accurately describes changes observed across different forms of life, but it does much more: It is the only physical process in our universe that can generate the objects we associate with life. This wider definition includes bacteria, cats, and trees, but also things such as rockets, mobile phones, and cities. None of these objects fluctuates into existence spontaneously, despite what popular accounts of modern physics may claim can happen. These objects are not random flukes. Instead, they all require a memory of the past to be made in the present. They must be produced over time—a time that continually moves forward. And yet, according to Newton, Einstein, Carnot, Boltzmann, and others, time is either nonexistent or merely emergent.

Reconciling Theories of Time

The times of physics and of evolution are incompatible. But this conflict has not always been obvious, because physics and evolution deal with different kinds of objects. Physics, particularly quantum mechanics, deals with simple and elementary objects: quarks, leptons, and force carrier particles of the Standard Model. Because these objects are considered simple, they do not require memory for the universe to make them (assuming sufficient energy and resources are available).

Assembly theory quantifies selection by making time a property of objects that could have emerged only via evolution.

Think of *memory* as a way to describe the recording of actions or processes that are needed to build a given object. When we get to the disciplines that engage with evolution, such as chemistry and biology, we find objects that are too complex to be produced in abundance instantaneously (even when energy and materials are available). They require memory, accumulated over time, to be produced. As Darwin understood, some objects can come into existence only through evolution and the selection of certain "recordings" from memory to make them.

This incompatibility of time across disciplines creates a set of problems that can be solved only by making a radical departure from the current ways that physics approaches time—especially if we want to explain life. Although current theories of quantum mechanics can explain certain features of molecules, such as their stability, they cannot explain the existence of DNA, proteins, RNA, or other large and complex molecules. Likewise, the second law of thermodynamics is said to give rise to the arrow of time and explanations of how organisms convert energy, but it does not explain the directionality of time, in which endless forms are built over evolutionary timescales with no final equilibrium or heat death for the biosphere in sight. Quantum mechanics and thermodynamics are necessary to explain some features of life, but they are not sufficient.

These and other problems led us to develop a new way of thinking about the physics of time, which we have called *assembly theory*. It describes how much memory must exist for a molecule or combination of molecules—the objects that life is made from—to come into existence. In assembly theory, this memory is measured across time as a feature of a molecule (or molecules) by focusing on the minimum memory required for that molecule to come into existence. Assembly theory quantifies selection by making time a property of objects that could have emerged only via evolution.

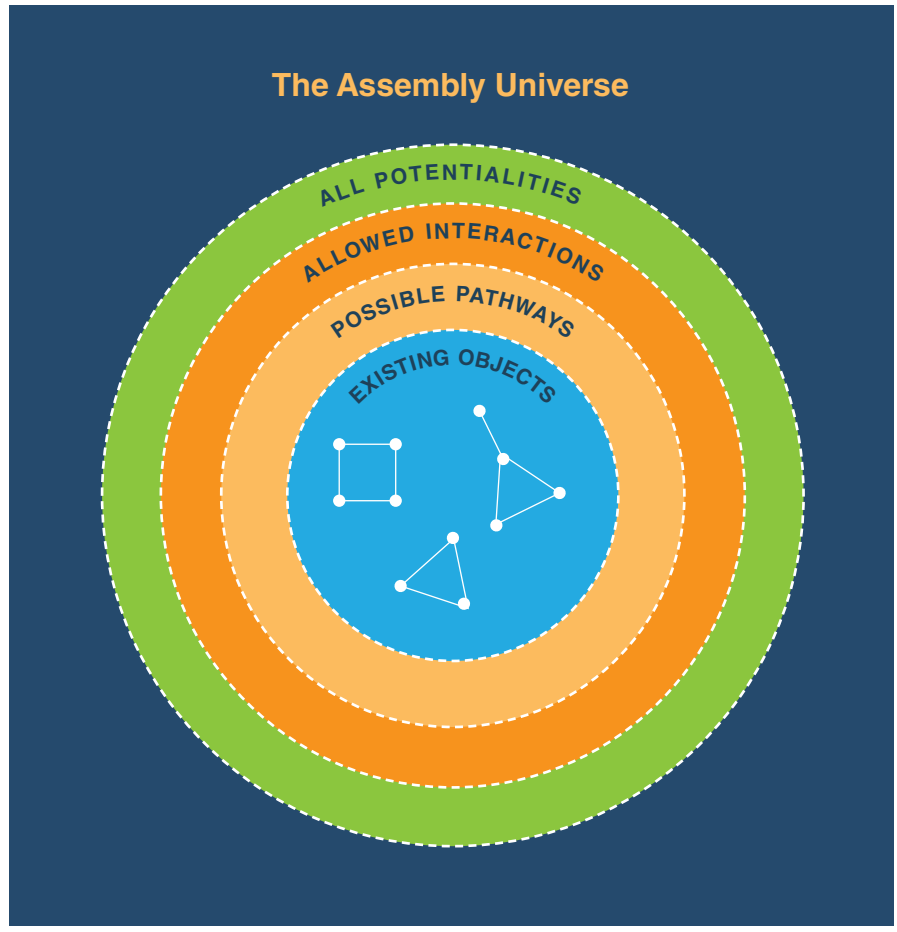
We began developing this new physics by considering how life emerges through chemical changes. The chemistry of life operates combinatorially as atoms bond to form molecules, and the possible combinations grow with each additional bond. These combinations are made from approximately 92 naturally occurring elements, which can lead to an

incredibly large number of molecules. Even when restricted to the six most-common elements found in life (carbon, hydrogen, nitrogen, oxygen, phosphorus, and sulfur) the possible combinations of unique molecules would be in excess of 10^{60}—that's 1 followed by 60 zeros. To become useful, each individual combination would need to be replicated billions of times—think of how many molecules are required to make even a single cell, let alone an insect or a person. Making copies of any complex object takes time, because each step required to assemble that object involves a search across the vastness of combinatorial space to select which molecules will take physical shape.

Consider the macromolecular proteins that living things use as catalysts within cells. These proteins are made from smaller molecular building blocks called amino acids, which combine to form long chains typically between 50 and 2,000 amino acids long. If every possible 100-amino-acid–long protein was assembled from the 20 most common amino acids that form proteins, the result would fill not just our universe but 10^{23} universes.

The space of all possible molecules is hard to fathom. As an analogy, consider the combinations you can build with a given set of Lego bricks. If the set contained only two bricks, the number of combinations would be small. However, if the set contained thousands of bricks, such as the 5,923-piece Lego model of the Taj Mahal, the number of possible combinations would be astronomical. If you specifically needed to build the Taj Mahal according to the instructions, the space of possibilities would be limited, but if you could build any Lego object with those 5,923 pieces, there would be a combinatorial explosion of possible structures that could be built—the possibilities grow exponentially with each additional brick you add. If you connected two Lego structures you had already built every second, you would not be able to exhaust all possible objects of the size of the Lego Taj Mahal set within the age of the universe.

Indeed, any space built combinatorially from even a few simple building blocks will have this property, including all possible cell-like objects built from chemistry, all possible organisms built from different cell types, all possible languages built from words or utterances, and all possible computer programs built from all possible instruction sets. The pattern here is that combinatorial spaces seem to show up when life exists. That is, life is evident when the space of possibilities is so large that the universe must select only some of that space to exist.



Barbara Aulicino, adapted from A. Sharma, et al. 2023

The fundamental particles of the universe can be combined in a staggering number of combinations, but assembly theory explains why only some of those combinations exist. In the "Assembly Universe," all particles (*green*) are first restricted by what interactions are possible according to the laws of physics (*orange*). The availability of resources and the "memory" of production pathways further narrows the possibilities (*peach*). And of those possibilities, even fewer are known to exist (*blue*).

Assembly theory is meant to formalize this idea. In assembly theory, objects are built combinatorially from other objects and, just as you might use a ruler to measure how big a given object is spatially, assembly theory provides a measure—called the *assembly index*—to measure how big an object is in time.

The Lego Taj Mahal set is equivalent to a complex molecule in this analogy. Reproducing a specific object, such as a Lego set, in a way that isn't random requires selection within the space of all possible objects. That is, at each stage of construction, specific objects or sets of objects must be selected from the vast number of possible combinations that could be built. Alongside selection, memory is also required: Information is needed in the objects that exist to assemble the specific new object, which is implemented as a sequence of steps that can be completed in finite time, like the instructions required to build the Lego Taj Mahal. More complex objects require more memory to come into existence.

In assembly theory, objects grow in their complexity over time through the process of selection. As objects become more complex, their unique parts will increase, which means *local memory* must also increase. Local memory is the causal chain of events in how the object is first "discovered" by selection and then created in multiple copies.

For example, in research into the origin of life, chemists study how molecules come together to become living organisms. For a chemical system to spontaneously emerge as life, it must self-replicate by forming, or catalyzing, self-sustaining networks of chemical re-



Gareth Milner/Wikimedia Commons/CC BY 2.0

The Lego Taj Mahal set includes 5,923 pieces, but only one combination of those pieces results in the intended structure. The chances of the Lego bricks falling in such a way as to build the Taj Mahal is near zero—human action is required to construct the pieces in the correct sequence. According to assembly theory, life itself is what enables the high density of information and local memory required to create unlikely, information-dense objects such as biological organisms.

actions. But how does the chemical system know which combinations to make?

We can see local memory in action in these networks of molecules that have "learned" to chemically bind together in certain ways. As the memory requirements increase, the probability that an object was produced by chance drops to zero, because the number of alternative combinations that weren't selected is just too high. An object, whether it's a Lego Taj Mahal or a network of molecules, can be produced and reproduced only with memory and a construction process. But memory is not everywhere; it is local in space and time. This means an object can be produced only where there is local memory that can guide the selection of which parts go where, and when.

In assembly theory, *selection* refers to what has emerged in the space of possible combinations. It is formally described through an object's copy number and complexity. *Copy number* or *concentration* is a concept used in chemistry and molecular biology that refers to how many copies of a molecule are present in a given volume of space. In assembly theory, *complexity* is as significant as the copy number. A highly complex molecule that exists only as a single copy is not important. What is of interest to assembly theory are complex molecules with a high copy number, which is an indication that the molecule has been produced by evolution. This complexity measurement is also known as an object's *assembly index*. This value is related to the amount of physical memory required to store the information to direct the assembly of an object and set a directionality in time from the simple to the complex. And, although the memory must exist in the environment to bring the object into existence, in assembly theory the memory is also an intrinsic physical feature of the object. In fact, it is the object.

Assembling Objects

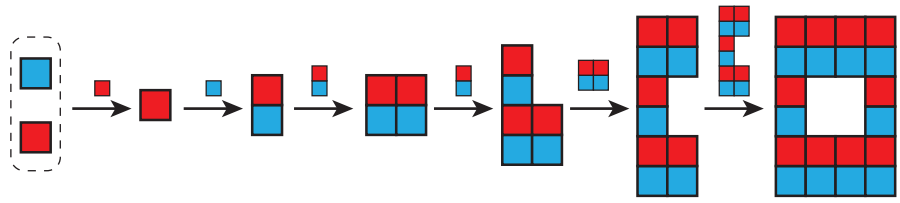
Life is stacks of objects building other objects that build other objects—it's objects building objects, all the way down. Some objects emerged only relatively recently, such as synthetic "forever chemicals" made from organofluorine chemical compounds. Others emerged billions of years ago, such as photosynthesizing cells. Different objects have different depths in time. And this depth is directly related to both an object's assembly index and copy number, which we can combine into a number: a quantity called *Assembly*, or A. The higher this Assembly number, the deeper an object is in time.

To measure Assembly in a laboratory, we chemically analyze an object to count how many copies of a given molecule it contains. We then infer the object's complexity, known as its *molecular assembly index*, by counting the number of parts it contains. These molecular parts, like the amino acids in a protein string, are often inferred by determining an object's molecular assembly index—a theoretical Assembly number. But we are not inferring theoretically. We are "counting" the molecular components of an object using three visualizing techniques: mass spectrometry, infrared, and nuclear magnetic resonance spectroscopy. Remarkably, the number of components we've counted in molecules maps to their theoretical Assembly numbers, which means we can measure an object's assembly index directly with standard lab equipment.

A high Assembly number—a high assembly index and a high copy number—indicates that a molecule can be reliably made by something in its environment. That "something" could be a cell that constructs high-Assembly molecules such as proteins, or a chemist who makes molecules with an even higher Assembly value, such as the

erectile dysfunction treatment drug Viagra. Complex objects with high copy numbers did not come into existence randomly but are the result of a process of evolution or selection. They are not formed by a series of chance encounters, but by selection in time. More specifically, a certain depth in time.

Assembly theory is a difficult concept. Even chemists find the idea hard to grasp because it is easy to imagine that complex molecules form by chance interactions with their environment. However, in the laboratory, chance interactions often lead to the production of *tar* rather than high-Assembly objects. Tar is a chemist's worst nightmare, a messy mixture of molecules that cannot be individually identified. It is frequently found in origin-of-life experiments. In the American chemist Stanley Miller's "prebiotic soup" experiment in 1953, the amino acids that formed at first turned into a mess of unidentifiable black gloop if the experiment was run too long (and no selection was imposed by the researchers to stop chemical changes taking place). The problem in these experiments is that the combinatorial space of possible molecules is so vast for high-Assembly objects that no specific molecules are produced in high abundance. Tar is the result.



S. M. Marshall, et al. 2021. CC 4.0

The construction of an object becomes more complex with the addition of more components. An object's Assembly value is determined by the fewest possible steps needed to make an object. Objects with more components or more complex pathways have higher Assembly values, and are more likely the result of a process of evolution or selection. Assembly theory can also help researchers determine whether complex objects found in space are signs of extraterrestrial life.

It's like throwing the 5,923 bricks from the Lego Taj Mahal set in the air and expecting them to come together, spontaneously, exactly as the instructions specify. Now imagine taking the pieces from 100 boxes of the same Lego set, throwing them into the air, and expecting 100 exact copies of the same building. The probabilities are incredibly low and might be zero, if assembly theory is on the right track. It is as likely as a smashed egg spontaneously reforming.

But what about complex objects that occur naturally without selection or evolution? What about snowflakes, minerals, and complex storm systems? Unlike objects generated by evolution and selection, these do not need to be explained through their depth in time. Although individually complex, they do not have a high Assembly value because they form randomly and require no memory to be produced. They have a low copy number because they never exist in identical copies. No two snowflakes are alike, and the same goes for minerals and storm systems.

Combinatorial spaces seem to show up when life exists. That is, life is evident when the space of possibilities is so large that the universe must select only some of that space to exist.

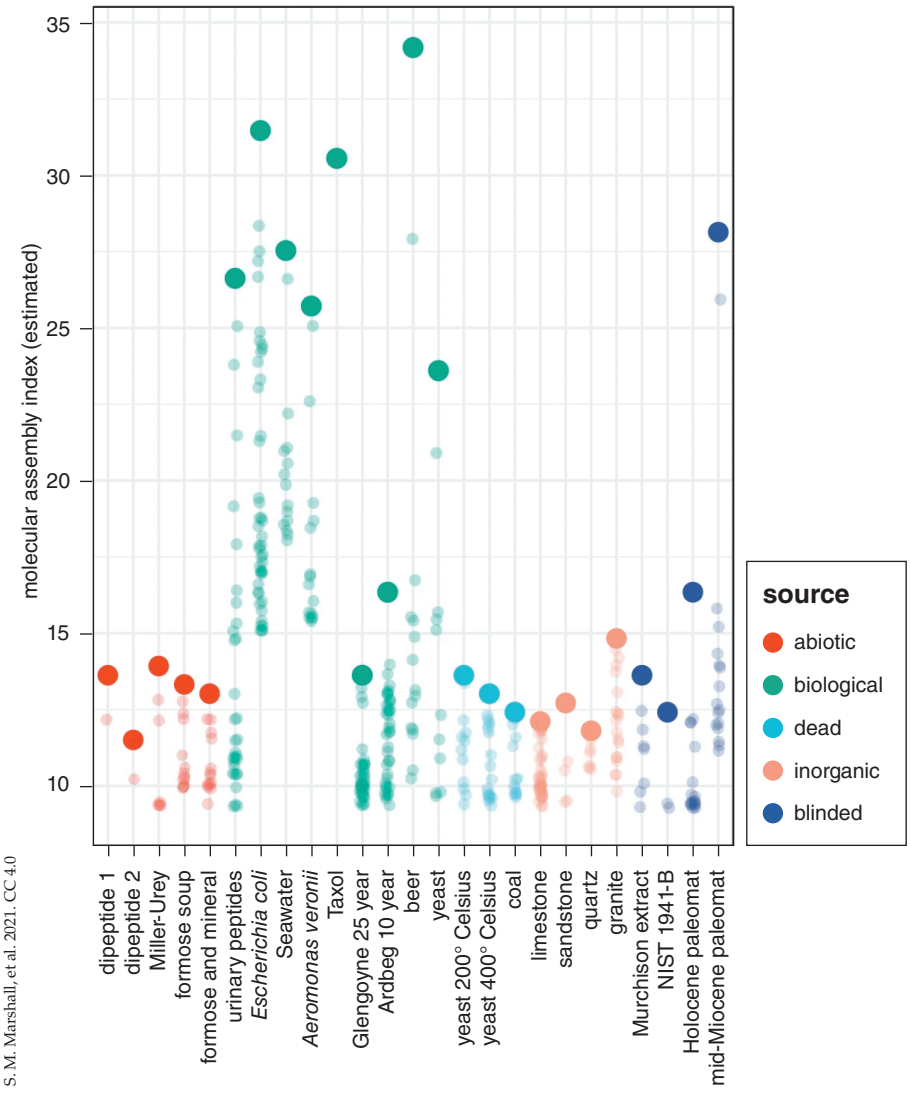
Signs of Life

Assembly theory not only changes how we think about time, but how we define life itself. By applying this approach to molecular systems, it should be possible to measure if a molecule was produced by an evolutionary process. That means we can determine which molecules could have been made only by a living process, even if that process involves chemistries different to those on Earth. In this way, assembly theory can function as a universal life-detection system that works by measuring the assembly indexes and copy numbers of molecules in living or nonliving samples.

In our laboratory experiments, we found that only living samples produce high-Assembly molecules. Our teams and collaborators have reproduced this finding using an analytical technique called *mass spectrometry*, in which molecules from a sample are "weighed" in an electromagnetic field and then smashed into pieces using energy. Smashing a molecule to bits allows us to measure its assembly index by counting the number of unique parts it contains. Through this technique, we can work out how many steps were required to produce a molecular object, and then quantify its depth in time with standard laboratory equipment.

To verify our theory that high-Assembly objects can be generated only by life, the next step involved testing living and nonliving samples. Our teams have been able to take samples of molecules from across the Solar System, including diverse living, fossilized, and abiotic systems on Earth. These solid samples of stone, bone, flesh, and other forms of matter were dissolved in a solvent and then analyzed with a high-resolution mass spectrometer that can identify the structure and properties of molecules. We found that only living systems produce abundant molecules, with an assembly index above an experimentally determined value of 15 steps. The cutoff between 13 and 15 is sharp, meaning that molecules made by random processes cannot get beyond 14 steps. We think this threshold is indicative of a phase transition where the physics of evolution and selection must take over from other forms of physics to explain how a molecule was formed.

These experiments verify that only objects with a sufficiently high Assembly number—highly complex and copied molecules—seem to be found in life. What is even more exciting is that we can find this information without knowing anything else about the molecule present. Assembly theory can determine whether molecules from anywhere in the universe were derived from evolution or not, even if we don't know what chemistry is being used.



S. M. Marshall, et al. 2021. CC 4.0

The authors estimated the *molecular assembly index* for a variety of substances to determine the relative complexity of the materials ("blinded" indicates that the samples were of unknown origin at the time of testing). Higher index numbers are uniquely associated with substances that are alive (such as yeast) or were created by a living organism (such as urine). These measurements can help to delineate what constitutes life, both on Earth and in outer space.

The possibility of detecting living systems elsewhere in the galaxy is exciting, but more exciting for us is the possibility of a new kind of physics, and a new explanation of life. As an empirical measure of objects uniquely producible by evolution, assembly theory unlocks a more general theory of life. If assembly theory holds, its most radical philosophical implication is that time exists as a material property of the complex objects created by evolution. That is, just as Einstein radicalized our notion of time by unifying it with space, assembly theory points to a radically new conception of time by unifying it with matter.

It is radical because, as we noted, time has never been fundamental in the history of physics. Newton and some quantum physicists view time as a backdrop. Einstein thought it was an illusion. And in the work of those studying thermodynamics, time is understood as merely an emergent property. Assembly theory treats time as fundamental and material: Time is the stuff out of which things in the universe are made. Objects created by selection and evolution can be formed only through the passing of time. But don't think about this time like the measured ticking of a clock or a sequence of calendar years. Time is a physical attribute. Think about it in terms of Assembly, a measurable intrinsic property of a molecule's depth or size in time.

Another reason this idea is radical is that it allows physics to explain evolutionary change. Physics has traditionally studied objects that the universe can spontaneously assemble, such as elementary particles or planets. Assembly theory, on the other hand, explains evolved objects, such as complex molecules, biospheres, and computers. These complex objects exist only along lineages where information has been acquired specific to their construction.

If we follow those lineages back, beyond the origin of life on Earth to the origin of the universe, it would be logical to suggest that the memory of the universe was lower in the past, which means that the universe's ability to generate high-Assembly objects is fundamentally limited by its size in time. Just as a semitrailer truck will not fit inside a standard home garage, some objects are too large in time to come into existence in intervals that are smaller than their assembly index. For complex objects such as computers to exist in our universe, many other objects needed to form first: stars, heavy elements, life, tools, technology, and the abstraction of computing. This process takes time and is critically path-dependent due to the causal contingency of each innovation made. The early universe may not have been capable of computation as we know it, simply because not enough history existed yet. Time had to pass and be materially instantiated through the selection of the computer's constituent objects. The same goes for Lego structures, large language models, new pharmaceutical drugs, the "technosphere," or any other complex object.

The consequences of objects having an intrinsic material depth in time is far reaching. In the block universe, everything is treated as static and existing all at once, which means that objects cannot be ordered by their depth in time, and selection and evolution cannot be used to explain why some objects exist and not others. Reconceptualizing time as a physical dimension of complex matter and setting a directionality for time could help us solve such questions. Making time material through assembly theory unifies several perplexing philosophical concepts related to life in one measurable framework. At the heart of this theory is the assembly index, which measures the complexity of an object. It is a quantifiable way of describing the evolutionary concept of selection by showing how many alternatives were excluded to yield a given object. Each

step in the assembly process of an object requires information, memory, to specify what should and shouldn't be added or changed. In building the Lego Taj Mahal, for example, we must take a specific sequence of steps, each directing us toward the final building. Each misstep is an error, and if we make too many errors we cannot build a recognizable structure. Copying an object requires information about the steps that were previously needed to produce similar objects.

Assembly theory can function as a universal life-detection system that works by measuring the assembly indexes and copy numbers of molecules in living or nonliving samples.

Assembly theory is a causal theory of physics because the underlying structure of an assembly space—the full range of required combinations—orders things in a chain of causation. Each step relies on a previously selected step, and each object relies on a previously selected object. If we removed any steps in an assembly pathway, the final object would not be produced. Buzzwords often associated with the physics of life, such as "complexity," "information," "memory," "causation," and "selection," are material because objects themselves encode the rules to help construct other complex objects. This process could be the case in mutual catalysis where objects reciprocally make each other. Thus, in assembly theory, time is essentially the same thing as information, memory, causation, and selection. They are all made physical because we assume they are features of the objects described in the theory, not the laws of how these objects behave. Assembly theory reintroduces an expanding, moving sense of time to physics by showing how its passing is the stuff complex objects are made of: The size of the future increases with complexity.

Time Is Fundamental

This new conception of time might solve many open problems in fundamental physics. The first and foremost is the debate between determinism and contingency. Einstein famously said that God "does not play dice," and many physicists are still forced to conclude that determinism holds, and our future is closed. But the idea that the initial conditions of the universe, or any process, determine the future has always been a problem. In assembly theory, the future is determined, but not until it happens. If what exists now determines the future, and what exists now is larger and more information-rich than in the past, then the possible futures also grow larger as objects become more complex, because there is more history existing in the present from which to assemble novel future states. Treating time as a material property of the objects it creates allows novelty to be generated in the future.

Novelty is critical for our understanding of life as a physical phenomenon. Our biosphere is an object that is at least 3.5 billion years old by the measure of clock time (Assembly is a different measure of time). But how did life get started? What allowed living systems to develop intelligence and consciousness? Traditional physics suggests that life "emerged." The concept of emergence captures how new structures seem to appear at higher levels of spatial organization that could not be predicted from lower levels. Examples include the wetness of water, which is not predicted from individual water molecules, or the way that living cells are made from individual nonliving atoms. However, the objects traditional physics considers emergent become fundamental in assembly theory. From this perspective, an object's "emergentness"—how far it departs from a physicist's expectations of elementary building blocks—depends on how deep it lies in time. This idea points us toward the origins of life, but we can also travel in the other direction.

If we are on the right track, assembly theory suggests time is fundamental. It suggests change is not measured by clocks but is encoded in chains of events that produce complex molecules with different depths in time. Assembled from local memory in the vastness of combinatorial space, these objects record the past, act in the present, and determine the future, which means the universe is expanding in time, not space—or perhaps space emerges from time, as many current proposals from quantum gravity suggest. Though the universe may be entirely deterministic, its expansion in time implies that the future cannot be fully predicted, even in principle. The future of the universe is more open-ended than we could have predicted.

Time may be an ever-moving fabric through which we experience things coming together and apart. But the fabric does more than move—it expands. When time is an object, the future is the size of the universe.

References

Doran, D., E. Clarke, G. Keenan, E. Carrick, C. Mathis, and L. Cronin. 2021. Exploring the sequence space of unknown oligomers and polymers. *Cell Reports Physical Science* 2:100685.

Liu, Y., C. Mathis, M. D. Bajczyk, S. M. Marshall, L. Wilbraham, and L. Cronin. 2021. Exploring and mapping chemical space with molecular assembly trees. *Science Advances* 7:eabj2465.

Marshall, S. M., et al. 2021. Identifying molecules as biosignatures with assembly theory and mass spectrometry. *Nature Communications* 12:3033.

Marshall, S. M., D. G. Moore, A. R. G. Murray, S. I. Walker, and L. Cronin. 2022. Formalising the pathways to life using assembly spaces. *Entropy* 24:884.

Sharma, A., D. Czégel, M. Lachmann, C. P. Kempes, S. I. Walker, and L. Cronin. 2023. Assembly theory explains and quantifies the emergence of selection and evolution, v3. arXiv:2206.02279.

Sara Walker is an astrobiologist and theoretical physicist at Arizona State University, where she is deputy director of the Beyond Center for Fundamental Concepts in Science and professor in the School of Earth and Space Exploration. She is also external professor at the Santa Fe Institute and a fellow at the Berggruen Institute. Lee Cronin is Regius Chair of Chemistry at the University of Glasgow in Scotland and CEO of Chemify. This article is adapted from one that appeared in Aeon, *aeon.co. Email for Walker: sara.i.walker@asu.edu; email for Cronin: lee.cronin@glasgow.ac.uk*

Scientists' Nightstand

The Scientists' Nightstand, American Scientist's *books section, offers reviews, review essays, brief excerpts, and more. For additional books coverage, please see our* Science Culture *blog channel online, which explores how science intersects with other areas of knowledge, entertainment, and society.*

ALSO IN THIS ISSUE

BETTER LIVING THROUGH BIRDING: Notes From a Black Man in the Natural World. By Christian Cooper.
page 312

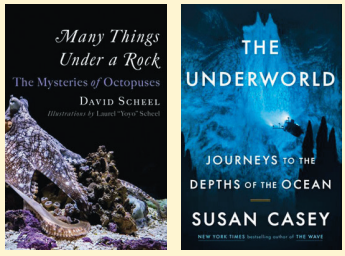
THE OCTOPUS IN THE PARKING GARAGE: A Call for Climate Resilience. By Rob Verchick.
page 315

ONLINE

On our *Science Culture* blog:
www.americanscientist.org/blogs/science-culture

Summer STEM Reads You Won't Want to Miss

Book Review Editor Jaime Herndon looks at recent releases covering oceanography, astronomy, biology, mathematics, and more.



An Ocean of Reasons to Map the Seafloor

Haley Cabaniss

THE DEEPEST MAP: The High-Stakes Race to Chart the World's Oceans. Laura Trethewey. 304 pp. Harper Wave, 2023. $32.00.

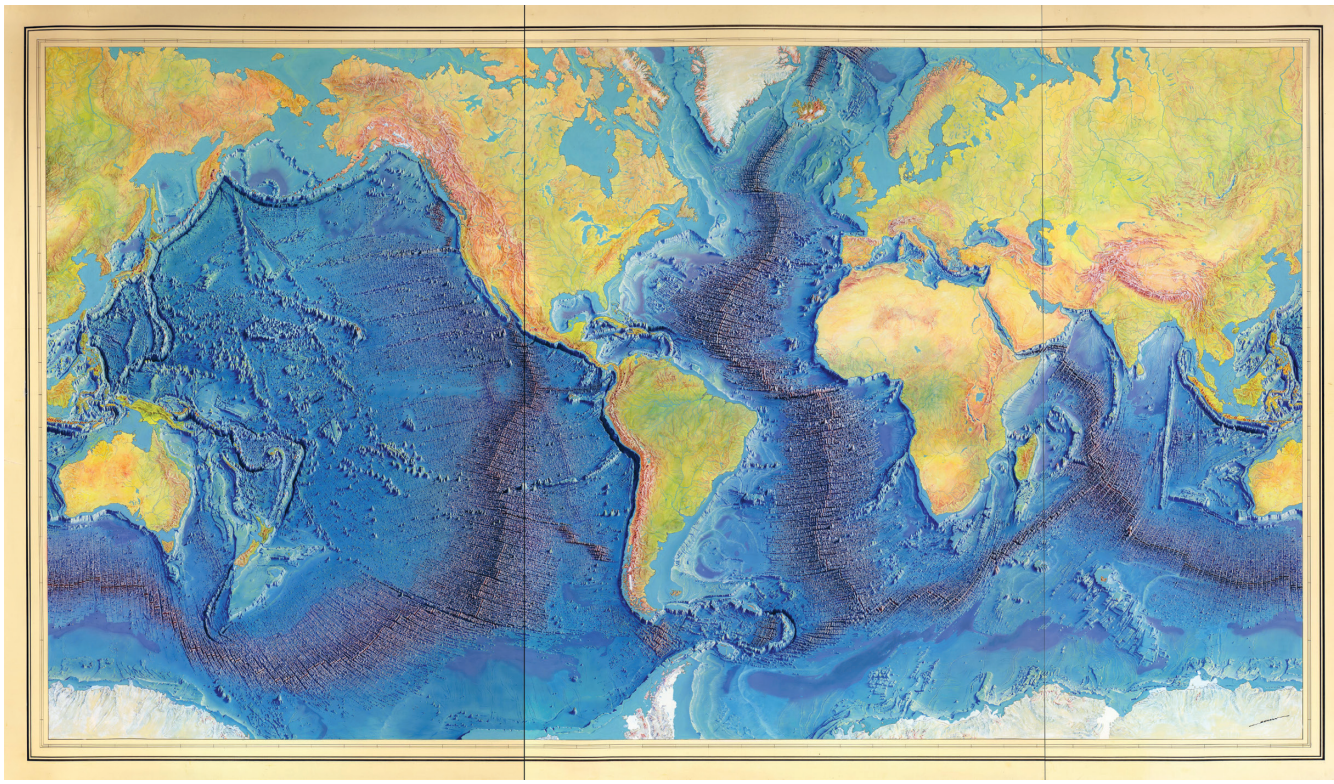
In her latest book, *The Deepest Map: The High-Stakes Race to Chart the World's Oceans*, award-winning environmental and ocean journalist Laura Trethewey documents the urgency to explore undiscovered ocean expanses and map the seafloor. Although the ocean covers more than 70 percent of our planet, just 25 percent of the Earth's seafloor is mapped. To describe the scope of this problem, the overused phrase "we know more about the surface of the Moon and Mars than we do about our own seafloor" is commonly tossed around ocean science circles. However, detailed maps of the Earth's seafloor, and changes in its shape, are critical for safe navigation, trade, and global supply chains. Seafloor maps provide key insight into marine hazards such as earthquakes, tsunamis, and volcanic eruptions. Such maps are also important for exploring seafloor resources that may fuel innovation in everything from electric vehicles to pharmaceutical industries.

Considering climate-driven oceanic changes, increased interest in mining the seafloor, and a growing population of more than 8 billion people who rely on healthy oceans, we need 100 percent of the Earth's seafloor mapped, and we need it now. The Nippon Foundation-GEBCO (General Bathymetric Chart of the Oceans) Seabed 2030 Project, typically referred to as Seabed 2030, is a current global initiative to map the entirety of Earth's seafloor by the year 2030. Trethewey explores the opportunity that Seabed 2030 and a global map of the Earth's seafloor provides, the arduous task of producing it during this decade, and some surprising consequences should it fail.

She details the Five Deeps expedition that took place in 2018 and 2019, during which private equity investor and world explorer Victor Vescovo became the first person to dive in a submersible to the deepest parts of all five oceans: the Atlantic, Pacific, Indian, Arctic, and Southern. To complete such an impressive feat required a support team of scientists and engineers, and, most importantly, a substantial mapping effort to determine the deepest location in each ocean basin.

If you Google "global seafloor map," you'll find dozens of images showing depth variation or *bathymetry* of the Earth's seafloor. Such results give the illusion of completion, of a comprehensive map of the global ocean basin. However, these images are not maps, but rather, predictions of seafloor depth, estimated from subtle changes in satellite measurements of gravity. Although satellite gravity measurements provide an excellent tool for predicting bathymetry, they do so at a resolution that is seldom practical for any real-world application, and certainly not for determining the deepest point in each of the world's oceans. Today, reliable seafloor maps are produced aboard survey vessels using sonar. Just as bats use sound for navigation through the process of echolocation, humans use sound to determine the distance from a specialized sonar system on a vessel to the surface of the Earth's seafloor.

Without mapping target locations for the deepest depths and determining their locations in each ocean basin, Vescovo's record was at risk of



Marie Tharp and Bruce Heezen published the first complete world map of the ocean floor in 1977. They did so with depth measurements from U.S. Navy submarines, and their work verified the theory of continental drift. This painting by Heinrich C. Berann was created the same year and accompanied the manuscript. From Library of Congress Geography and Map Division.

being surpassed by future ocean explorers who might have access to better bathymetric maps. He needed a *hydrographer*—someone who measures ocean depths and maps underwater topography—to map the seafloor, plan his dive sites, and validate his records. That's where Cassie Bongiovanni, a recent graduate with a master's degree in ocean mapping, came in. During the roughly yearlong Five Deeps expedition, Bongiovanni mapped an area half the size of Brazil and determined the deepest locations in all the world's oceans: 8,376 meters in the Atlantic Ocean, 7,434 meters in the Southern Ocean, 7,192 meters in the Indian Ocean, 5,550 meters in the Arctic Ocean, and 10,925 meters at the Challenger Deep in the Pacific Ocean—the deepest place on Earth. Not only did Bongiovanni's work enable Vescovo to become the first person to verifiably descend to the deepest parts of each ocean, she also contributed significant new maps of previously unexplored regions to the Seabed 2030 initiative.

Mapping the entirety of the seafloor is no small feat. The oceans are unfathomably large, remote, and often hostile to the hydrographers responsible for mapping them. Such an initiative requires extraordinary resources: time, money, vessels, survey equipment, and specialists. Even when such resources are available, conditions at sea often lead to lost time and subpar data. Trethewey reveals widespread concern about the task of mapping the entire ocean basin by the end of the decade, and explores potential catalysts for success, such as high-tech unmanned surface vessels (USVs) that are capable of mapping large swaths of the seafloor with limited human intervention, or crowdsourcing maps in remote regions, where local communities may benefit the most from a novel map of their coastal waters. Achieving a global map of the seafloor by the year 2030 is a situation requiring collaboration, international cooperation, and every available resource. The task is daunting and the time line short, but the consequences of mapping at a comfortable pace are immense.

This urgency is apparent when Trethewey travels to the remote Inuit hamlet of Arviat, on the mainland of Nunavut, where people are strongly reliant on coastal waters for food. Although sea level is rising in many places around the globe, here the land is moving upward in a process known as *glacial rebound*, and as a result, sea level has the appearance of falling. Inuit people have relied on elders in the community for navigation advice in coastal waters, but with recent *shoaling*, or changes in the shape and behavior of the waves as the water depth decreases, community leaders are no longer able to confidently determine the shape of the seafloor and help vessels safely navigate. Af-

The task is daunting and the time line short, but the consequences of mapping at a comfortable pace are immense.

ter several fatal boating accidents, the community of Nunavut requested mapping support and are now being trained to map and monitor their own waters to avoid future preventable wrecks. In another example, Trethewey participates in archaeological excavations in Florida's coastal waters, where sea level rise has led to land loss and flooded cultural sites. As the sea level continues to rise, archaeologists are concerned that more sites will be lost. Without improved maps, these sites may be lost forever.

Although mapping the seafloor has historically been a project undertaken by academics and community leaders such as those Trethewey visited, there are larger, more powerful groups with special interest in seafloor mapping for reasons that include everything from industrial exploitation to national defense. Trethewey attends meetings of the International Seabed Authority and the GEBCO Sub-Committee on Undersea Feature Names (SCUFN), where she sees firsthand how nations can improve their political position by mapping their coastal waters, and by naming and claiming the seafloor. A nation's exclusive economic zone, over which they have jurisdiction for exploration and exploitation, extends 200 nautical miles (about 370.4 kilometers) off the coast. Improved maps have the potential to help expand an exclusive economic zone, and thus a country's borders. Even naming a newly mapped feature on the seafloor in international waters has the potential to help a nation extend their political reach. Nothing about the oceans is apolitical, adding yet another layer to scientific exploration.

In *The Deepest Map*, Trethewey braids her personal experiences and observations from joining expeditions of uncharted waters with reportage on and interviews with various stakeholders in ocean mapping, as well as Bongiovanni's journey mapping the deepest extremes of the world's oceans for Vescovo. She thoughtfully untangles the often overlooked and poorly understood needs for expanded seafloor maps, and makes the case that mapping the seafloor is indeed a high-stakes race against climate change, mining, and the reach of human influence. The book provides a glimpse into one of science's and society's greatest challenges—mapping the entirety of the Earth's seafloor—from a diversity of perspectives that highlight the complexity and necessity of this task, in a way that is accessible and interesting to both scientists and lay readers.

Haley Cabaniss is an assistant professor of marine geology at the College of Charleston (CofC), where she studies submarine volcanism and also helps to train the next generation of hydrographers through the CofC BEnthic Acoustic Mapping and Sonar (BEAMS) program.

Birding as a Social Lesson

Tenijah Hamilton

BETTER LIVING THROUGH BIRDING: Notes from a Black Man in the Natural World. Christian Cooper. 304 pp. Random House, 2023. $28.00.

When I became a birder during the summer of 2020, it was as though a veil had lifted from the mundanity of my pandemic-limited home life: open laptop, jiggle mouse, doom-scroll, close laptop, detach, despair. With birding, suddenly I was part of an immersive, sensory-rich world in which I could hear the birds sing louder than I ever had before. Although I was initially drawn to it as an escape, becoming a birder helped me feel more deeply embodied, present, and attuned to the natural world and myself. It makes sense to me that birder, comics writer, and explorer Christian Cooper's memoir *Better Living Through Birding: Notes from a Black Man in the Natural World* likewise examines the broader life experiences associated with his own practice of birding, because they are inextricable from one another.

Although this memoir is primarily personal, it is a story viewed through the lens of the natural world. Cooper introduces readers to his passion for birding, his childhood, adolescence, and early adulthood, as well as his experiences of being a queer Black man in the United States. He shares his sometimes strained relationships with family members, coming out to his family, his love of travel, and the ways his life changed during the early days of the COVID-19 pandemic. All of these ideas are woven in alongside "Birding Tips," which ostensibly contain practical guidance about birding but which also function as subtle commentaries on what he's writing about at that moment.

The book begins with a chapter titled "An Incident in Central Park." You may think you already know the incident in question: a viral encounter during the summer of 2020 that saw Cooper harassed by a white woman who was upset that he insisted that she leash her dog; but this is not that—not yet. No, this chapter is about Cooper racing out of his office to spot an elusive Kirtland's warbler. This part of the book is also where Cooper begins to share an interspersed set of commentaries he calls the "Seven Pleasures of Birding." Number three is The Joy of Scientific Discovery, in which he recounts his encounter with an aggressive bunting (a songbird related to the finch). Cooper had encroached on the bird, and fearing for its family's safety, the bunting was rightfully angry. Although fighting the man was a nonstarter, the bird *could* fight a nearby stick, and in an instance of transference, it did so with vigor. This experience inspired a deeper exploration into Cooper's own emotions, and specifically, the anger he would channel through a dark character he once used as an avatar in his late high school and college writing called Dr. Яamus. In Cooper's recounting, art imitates birding, which imitates life, proving again just how intertwined those experiences can be.

Like many people, it took Cooper a while to see that his intersecting identities and his varied experiences, all of which defied easy explanation, were not flaws but strengths. Early in the book, he writes about his "spark bird" and connects it to his undefinable identity. In the birding world, a spark bird is the first bird that instigates your love of the hobby. For Cooper, it was the common yet beguiling red-winged blackbird. The bird has been taxonomically controversial since the arrival of European colonists to North America, who wrongfully assumed it to be the same as the Eurasian blackbird. In fact, the red-winged blackbird is an icterid—part of a large family of birds that couldn't be more American than if they raided the star-spangled flip-flop bin at Old Navy on the Fourth of July.

The Eurasian blackbird is part of the genus *Turdus*, or thrush, which adds

ASTRONOMY IS FULL OF UNRESOLVED ENIGMAS

Lawyer & Judges Publishing Company, Inc.

RECENT DISCOVERIES IN ASTRONOMY a new Physics-based book for scientists who wish to explore new solutions to such problems as the origin of our solar system, the Sun's million-degree corona temperature and the mysterious center of our galaxy. One of the many surprises is a unique chapter on gravity and the discovery of a planetary nebula surrounding our solar system.

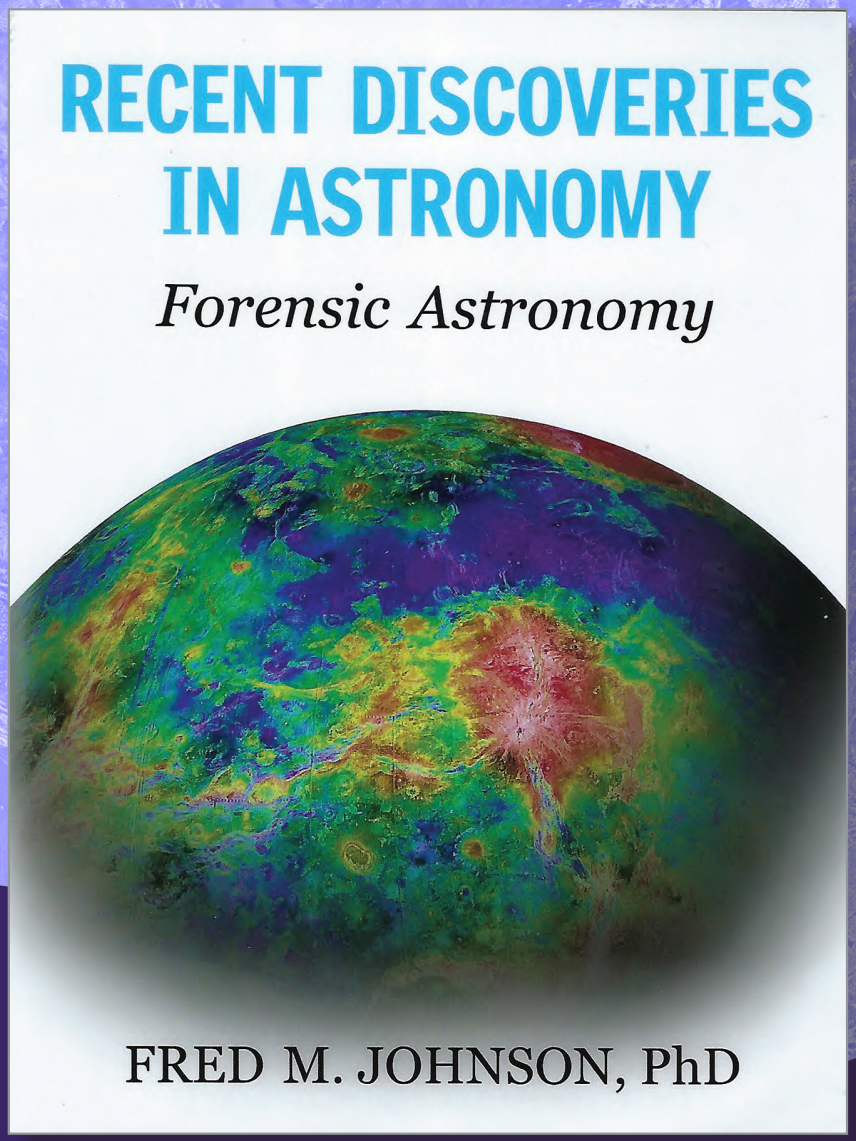


Image Credit: NASA/JPL/USGS

BUY NOW

www.lawyersandjudges.com
Product Code #6084



Andrew Maas/NYC Audubon

Christian Cooper takes viewers into the world of birds on his new National Geographic show *Extraordinary Birder*. Here, he is helping New York City Audubon scientists band a herring gull.

another layer of complexity. Although the *Turdus* genus has its own variation of redwing, the red-winged blackbird is not it. Cooper relates to the birder's fatigue at the mislabeling of the red-winged blackbird, recognizing it as similar to his own frustration with society's labels:

> I am no exception. Like everyone else, I had to sort through aspects of my own identity and where I fit in the social taxonomy, which labels fit and which labels chafed, and how the world may have misidentified me and pegged my kind all wrong. I had to grow comfortable in my own Black skin in a white world, in my own rainbow-queer body in an era where sexuality was only seen as black and white.

Growing up on Long Island, in a family with a bent toward science and civil rights activism, Cooper became a student of the world, learning in equal measure from the birding field guides and comic books he always kept in tow. After moving to Boston to attend Harvard University, Cooper began to come to terms with his sexuality, joining the Gay Students Association and confiding in his ragtag team of roommates, who rallied around him. Acceptance and a keenly developed sense of self-actualization propelled him toward a dream career in comic book writing, challenging the status quo in legacy brands such as Marvel, where he introduced the very first gay and lesbian characters.

Christian Cooper's love for birding resounds throughout the book, positioning it as a radical act of resistance and reclamation of joy from those who wish to steal it away.

It's not until the late second half of the book that Cooper ambles up to the life-altering moment in Central Park, the one that catapulted him to international notoriety. While birding in a section of Central Park known as the Ramble, he films his experience of being accosted by an increasingly agitated white woman who attempts to summon the police as a means of racialized intimidation, by calling 911 and reporting that a Black man is threatening her life. Although the claims are patently false (as proven by the video), the woman—who happens to share the surname Cooper—goes from annoyed during the confrontation to theatrical screams, yelps, and sobs when phoning 911.

For Cooper, as for many Black people, it is impossible to exist without at least subconsciously acknowledging that many white people see you as a threat simply for existing. Their supposition of fear can easily become an actual threat to your existence. This was the case for Ahmaud Arbery who, just a few months prior to Cooper's experience, was out for a jog in Georgia when he was targeted, chased down, and executed by two white men, who were not even charged with a crime for more than two months. The message was clear: You cannot assume you are safe while you are simply living your daily life. Even in the birding world, the sanctuary that Cooper had been part of for so long, the danger still threatened him.

These experiences are not tangential to the experience of Black birders; they are fundamental facets of it. Speaking for myself as a Black birder, I can only try to imagine the deep feeling of being at one with nature that my white peers seem to feel when out in nature, but that I may never tap into. A complete surrendering to the experience where they don't have to be simultaneously hyperaware of how they're dressed (definitely no hoodies), what they're carrying (Amy Cooper told police that Christian Cooper seemed threatening because he was holding a bicycle helmet), and above all, performing joviality. God forbid someone mistakes you at any time for being sketchy, suspicious, or—most dangerous of all—not belonging.

Author and anthropologist Zora Neale Hurston once wrote, "If you are silent about your pain, they will kill you and say you enjoyed it." By releasing the video of the incident in Central Park, Cooper demanded that the world bear witness to his pain,

and the response was powerful. Marginalized people heard the rallying call, and in response, they built a community of resplendent care, embodied in Black Birders Week. That is the ultimate legacy of that encounter, and it is fortified again in this book. Christian Cooper's love for birding resounds throughout the book, positioning it as a radical act of resistance and reclamation of joy from those who wish to steal it away.

Cooper grounds us in the facts of his life—his childhood curiosity about nature, his journey toward embracing his queer identity, the precarity and power of his racialized "otherness" in prestigious institutions such as Harvard and the Audubon Society—while also giving the reader permission to spread their own wings, guiding them into the birding world through tips and quippy anecdotes from his time in the field. He repeatedly illustrates how birding has given him a new lens through which to view his life and the intersections of his various identities. Cooper is a victor and not a victim; he is a man and no one's martyr. *Better Living Through Birding* gives a new generation of birders permission to define themselves for themselves, from this point forward.

Tenijah Hamilton is a writer, as well as an executive and creative producer, with a passion for using the transformative art of storytelling to instigate change. On her BirdNote podcast Bring Birds Back, *she talks with bird enthusiasts from various backgrounds, identities, and communities to figure out ways we can all work to save our feathered friends.*

Climate Change Is Here: What Now?

Yarrow Axford

THE OCTOPUS IN THE PARKING GARAGE: A Call for Climate Resilience. Rob Verchick. 288 pp. Columbia University Press, 2023. $32.00.

Climate change is no longer a distant threat: Supercharged wildfires, floods, and heat waves are already reshaping our world. Climate-related challenges that once seemed hypothetical now confront us in our own backyards. Rob Verchick's *The Octopus in the Parking Garage* is a book written for exactly this moment.

The octopus in the book's title became a social media celebrity in 2016, when one of Florida's increasingly common coastal floods ushered it alive into the bowels of a Miami parking garage. Photographed with its tentacles splayed in the garage's floodwaters alongside marooned SUVs, the octopus became a tweetworthy symbol that climate change had arrived. It gave visual urgency to the warnings that researchers had been issuing for years.

To many scientists (at least until recently), emphasizing climate resilience has felt like throwing in the towel.

In my more than two decades as a climate scientist, I have devoted my career to understanding the causes and consequences of climate change in the rapidly warming Arctic. Arguably, there is no way to study arctic climate change, watching atmospheric carbon dioxide rise unfettered all the while, without despairing at our failure to curb fossil fuel burning. Drastic reductions in greenhouse gas emissions are needed now more than ever, but Verchick makes a strong case for urgently prioritizing something else in parallel: climate resilience. In the very practical work of climate resilience, which requires that communities work together on local solutions, we may also find a cure for some of our despair.

Climate resilience, in Verchick's words, is "the capacity to manage and recover from a climate impact in a way that preserves a community's central character—the parts of its history, culture and economy that nourish the soul." Verchick, a professor of climate and environmental law who worked on climate adaptation policy for the U.S. Environmental Protection Agency (EPA) under the Obama Administration, writes from a position of extensive experience. Refreshingly, he also writes as someone eager to learn from the tribal leaders, community



François Libert/CC BY-NC-SA 2.0

This photo shows the Caribbean reef octopus (*Octopus briareus*), hunting at night for food. Another octopus of this same species famously appeared in a Florida parking garage in 2016. It was suspected that the octopus was in a drainage pipe that had become submerged due to rising sea levels, and with a "supermoon" high tide, the storm drain reversed, causing the octopus to be propelled up into the garage.

organizers, and concerned residents who know the most about their own local climate challenges. This book incorporates a wide range of voices and attempts to grapple with the almost infinite complexity of real-world climate adaptation.

The first section of the book makes a general case for why climate resilience should be prioritized alongside urgent work to slow climate change. Verchick argues for "moonshot" level funding for climate adaptation, and he describes in general terms what climate-resilient adaptations should look like. These chapters wax philosophical but they provide an important counterpoint to a long history in which most of our public attention has been captured by the need to reduce emissions. To many scientists (at least until recently), emphasizing climate adaptation has felt like throwing in the towel. But in Verchick's words, "Climate resilience is not about giving up or giving ground. It's about getting real."

The octopus in the parking garage was telling us it was time for action, everywhere.

The early chapters of *The Octopus in the Parking Garage* also help wrap readers' minds around the complexities of climate adaptation, before the next section presents detailed examples. They provide a theoretical framework for thinking about justice, trade-offs, and the difficult decisions and inevitable losses that are to come. In the United States and globally, climate change will hit people of color and people living in poverty the hardest. A key take-home is that it is essential to center social justice in planning for climate resilience, and that historically disadvantaged groups must have strong voices and political power throughout the process if genuine, shared resilience is to be achieved.

I especially enjoyed the second, longer part of the book, which describes specific climate resilience challenges and how individuals, communities, and government agencies have attempted to meet them. Here, Verchick takes us on an expansive journey across the United States. He visits Louisiana's coastal swamps, where epic efforts at resilience are underway but where tribal communities still face the possibility of losing the very ground they live upon to a combination of sinking local land and rising seas. We learn how the disastrous Hurricane Sandy and some very determined lawyers forced New York's electric utility company to modernize in forward-looking preparation for even bigger future floods. Verchick dives to a coral reef and explores the complexity of wildfire management in the western United States. We visit the Mojave Desert, where it is debated what steps federal land managers should take to secure the future of the iconic Joshua tree. Verchick consults with locals who are working on the front lines of each of these challenges, and gives wide-ranging examples of the forms resilience can take—from changes in state and city governance to hands-on species conservation work to cultural change. Parts of this section are extremely detailed, but the diversity of communities, climate challenges, leaders, and solutions that Verchick features make these stories both fascinating and educational.

These chapters pinpoint diverse actions each of us can take to help the communities and landscapes we care about to maintain their best qualities in the face of climate change. Verchick shares some fresh ideas for engaging with local organizations that are already working on the front lines of climate resilience—for example, by visiting your favorite state or national parks and expressing your concerns about climate impacts to park officials, or attending city housing policy meetings and speaking up about wildfire risks. Kids can educate family members and present at local hearings.

Verchick also argues that focusing on resilience can "clear a path for more abstract discussions on containing carbon emissions," perhaps opening skeptical minds more than scientific data ever will. In dealing with "local, immediate, and concrete" problems that relate to climate change, individuals and organizations can arrive organically at an awareness of the threats posed by climate change to the things they value. For example, the politically divided and fossil fuel–rich state of Louisiana united around policies aimed at "saving [its] coasts from annihilation," and that ultimately led to quite logical, if seemingly improbable, support for policies to reduce greenhouse gas emissions.

By design, this book focuses on the United States. Verchick offers not a radical reimagining of America, so much as a vision for how Americans could work largely within existing frameworks of governance to improve resilience and—a topic I am glad to see woven throughout—to center justice for the traditionally underserved communities that will be most affected by climate change. The author fully acknowledges that communities' abilities to adapt will be limited by funding priorities and other policy decisions of state and federal governments, which is also an acknowledgment that we must elect leaders at every level who will prioritize addressing climate change.

The Octopus in the Parking Garage brings to life how climate change is affecting people all over the United States and what coping with ongoing climate change will look like. It sparkles with colorful examples and even hope. Each of us can take practical steps to preserve the places and things we love most. From policymakers to worried parents like me, anyone concerned about climate change will find ideas here, as well as incentives to take action. Above all, this book is an urgent call to give climate resilience the monumental attention it requires in this pivotal moment.

That octopus in the parking garage was telling us it was time for action, everywhere. Seven years after the stranded cephalopod was rescued and splashed back into the sea, it is not too late to slow climate change or to make our world more resilient to the changes we have already caused—but the need to do so becomes more dire with every record-breaking fire season, heat wave, and flood we endure.

Yarrow Axford is a professor of earth and planetary sciences at Northwestern University. Her field- and lab-based research investigates past climate change in the Arctic, with a recent focus on Greenland (Kalaallit Nunaat to Greenlanders).

September–October 2023 Volume 32 Number 05

Sigma Xi Today

A NEWSLETTER OF SIGMA XI, THE SCIENTIFIC RESEARCH HONOR SOCIETY

Important Upcoming Dates for Sigma Xi

Fall Nominations
Who will you nominate for Sigma Xi membership this fall? Nominations require two nominators. Sigma Xi staff can help you find a second nominator if you need one. For more information, visit **sigmaxi.org/nominate** or email **membership@sigmaxi.org**.

October 1: Fall deadline for Grants in Aid of Research (GIAR) applications
New award amounts are now available for **Sigma Xi Members**! Graduate student members are now eligible for GIAR awards up to $5,000, and undergraduate student members are eligible for amounts up to $2,000. Nonmembers, both graduate and undergraduate, are still encouraged to apply for award amounts up to $1,000. Visit **sigmaxi.org/giar** for more information.

October 29: Last chance for IFoRE discount rates
Don't miss out on discounts! All conference registrations beyond this date will be at full registration rates. Visit **experienceIFoRE.org/attend** for more information.

November 9: Sigma Xi Annual Business Meeting & Assembly of Delegates —Long Beach, California

November 10–12: International Forum on Research Excellence (IFoRE) — Long Beach, California

Sigma Xi Today is managed by Jason Papagan and designed by Chao Hui Tu.

From the President

Scientists as Dream-Sowers



Paul was a particularly bright student. On the day of the defense of his doctoral dissertation, just before his committee started deliberations, I opened questions to the public. I am confident that my students are sufficiently prepared to answer any question. But then Paul was tripped up by a question concerning how long it would take for tissue-transillumination with near-infrared light—the subject of his thesis—to solve the cancer problem.

Ten years later, I was giving an award to another bright young scientist who had developed a technique for generating near-infrared light to image cancerous cells in the brain, an approach that he promised would replace drilling holes in the skull to image tumors. I was in shock when I heard this researcher overpromise the potential of his technique. The researcher in the latter case had obviously told the same promising story to his funding source as he had written in the published paper. But if mouse skull was actually there during his experiment, imaging with infrared light would not be possible.

I tell these stories to illustrate one of the reasons why there now seem to be so many people who no longer trust science. This erosion of trust in science was most obvious during the COVID-19 pandemic, resulting in an ever-increasing chasm between those who trust science blindly and those who are skeptical about its promises, potential, and actual accomplishments. Who tells the stories that, at least to some people, carry little credibility?

The general progress of medical research has been extremely slow. Fifty years ago, President Nixon promised that cancer would be cured before the end of the century, just as President Kennedy had promised to put a human on the Moon by the end of the 1960s. But space travel proved a simpler puzzle than progress in health science. Today, one in two people will experience a brush with cancer, compared with one in three only 20 years ago.

When President Kennedy first talked about conquering space, the most promising medication against several cancers, including the extremely stealthy ovarian cancer, was being discovered. In its less harsh form, doxyl was not generally available 40 years later because of incredible difficulties in process control. While young medical and scientific researchers were promising hyped-up breakthroughs in cancer research, exasperated cancer patients pleaded with the government to streamline the production of an existing, proven, and effective medication, doxyl, a so-called last defense. Finally, the U.S. government designated it an orphan drug, and assured its production.

May scientists never forget that, while they have the current skills and ideas to forge scientific progress, humans—in their own heads and in dusty libraries they keep—store all the knowledge that the world acquired through the ages.

Marija Strojnik

Marija Strojnik

Nobel Laureate Frances Arnold Highlights Keynote Lineup



Sigma Xi, The Scientific Research Honor Society is proud to announce the Keynote Speakers for its 2023 International Forum on Research Excellence (IFoRE). Dr. Frances Arnold, the Linus Pauling Professor of Chemical Engineering, Bioengineering, and Biochemistry at the California Institute of Technology, will be the featured Saturday speaker for the annual conference, which brings together STEM students and professionals from all scientific disciplines to celebrate excellence in the research enterprise and its impact on the betterment of society. Other Keynote Speakers include Worcester Polytechnic Institute's Suzanne Scarlata, Florida Atlantic University's James M. Sullivan, Northern Arizona University's Miguel José Yacamán, and McGill University's Richard Boudreault.

IFoRE '23 is an international event taking place November 10–12 in Long Beach, California. In addition to the lineup of distinguished speakers, the conference will include panel discussions, workshops, student research awards, networking opportunities, and social events. IFoRE is open to all students and professionals across the research enterprise and will follow the one-day annual business meeting and Assembly of Delegates for Sigma Xi chapter leaders on November 9.

Dr. Arnold received the 2018 Nobel Prize in Chemistry for pioneering directed evolution methods used to make enzymes for applications in sustainable chemistry across medicine, consumer products, agriculture, fuels, and chemicals. In 2021, she was appointed co-chair of the President's Council of Advisors on Science and Technology (PCAST) by President Biden. She also received the National Medal of Technology and Innovation from President Obama. Her Key Thought session at IFoRE will discuss evolution and clean chemistry, including a special breakout discussion for IFoRE student attendees to connect and chat with the Nobel Laureate.

Learn more at **experienceIFoRE.org**

Register by October 29 for Early Bird Savings

Discounts available for Sigma Xi members and students!

STEM Art & Film at IFoRE

Sigma Xi has a celebrated tradition of showcasing STEM-themed art at its annual conference. The convergence of science and art often represents complementary mindsets that fortify the outcomes of science, theory, and design. Artists who submit entries to IFoRE will have the opportunity to display their art in a gallery-like setting at the IFoRE exhibit hall, sell their art to interested attendees, and compete to win monetary prizes at the IFoRE STEM Art & Film Awards.

Join us to explore science, technology, engineering, and math (STEM) through different forms of visual and performing arts, including classic fine art, photography, painting, 3D visualization, film, and live performance. Sigma Xi invites all interested artists to submit their work, and full conference attendees may include art submissions as part of their paid registration.

To learn more and submit your work, visit **experienceIFoRE.org/stem-art**.



Winning artwork from past conferences

FACES of GIAR

Emma Dougherty

Grant: $1,000 in fall 2021
Education level at time of the grant: Undergraduate student



Project Description: Artificial light at night (ALAN) has been shown to impact the behaviors and population dynamics of wildlife, often with negative consequences. The chimney swift (*Chaetura pelagica*) is a small aerial insectivore distributed across the Americas, utilizing artificial structures such as chimneys for nesting, breeding, and roosting. Once abundant across its range, the chimney swift population has been steadily declining and is currently listed as "vulnerable" by the International Union for Conservation of Nature (IUCN). We hypothesized that ALAN would affect the roosting behavior of chimney swifts, and specifically that higher levels of ALAN would cause delayed entry into their roosting sites. We assessed ALAN using a sky quality meter that this grant allowed us to purchase. Furthermore, video cameras and trap cameras were purchased using this grant, allowing us to track the birds' behavior for analysis. Chimney swifts have been shown to play an essential role in ecosystem services by helping to control the insect population, and by providing a potential food source for raptors. The findings of this experiment shed light on the causes of chimney swift population decline and helped us to understand the broader effects of ALAN on roosting behaviors in wildlife populations.

How did the grant process or the project itself influence you as a scientist/researcher? Conducting this research allowed me to gain the ability to develop hypotheses, design and conduct experiments, and interpret results. Additionally, my scientific communication skills, both oral and written, were strengthened during this project. I was able to present my preliminary findings, which helped me learn how to communicate effectively about science to both expert and general audiences.

What is one piece of advice or tip you would give to future grant applicants? To future grant applicants who are just beginning their research, I would say that before you start collecting data for analysis, read as many published papers about similar research as you can! This will help you find the best techniques for data collection and allow you to troubleshoot any potential issues before they arise.

GRANTS IN AID OF RESEARCH
SIGMA XI

Spring 2023 GIAR Awards

Sigma Xi awarded 108 student research grants for the spring 2023 cycle of its Grants in Aid of Research (GIAR) program. Since 1922, the Society's GIAR program has been funding research for undergraduate and graduate students, and currently awards grants biannually in the fall and spring.

This year's Committee on Grants in Aid of Research, along with a panel of guest reviewers, evaluated 780 applications across most research disciplines. Chaired by Drew Coleman of the University of North Carolina at Chapel Hill, the committee awarded grants to 16 undergraduate students, 28 master's students, and 64 doctoral students. Grant amounts ranged from $500 to $4,790, and a total of $125,631 was awarded.

Visit **sigmaxi.org/GIAR-recipients** to view the names and research projects of the spring 2023 awardees.

Visit **sigmaxi.org/GIAR** to learn more about the program and its recent centennial celebration, read stories from past recipients, and submit applications for future grants. The deadline for fall grant applications is October 1, 2023.

Students Awarded for Team Science at 2023 Regeneron International Science and Engineering Fair





Each year, more than 1,500 ninth through twelfth graders worldwide are selected to compete at the Regeneron International Science and Engineering Fair (Regeneron ISEF) by winning a top prize at a local, regional, state, or national science fair affiliated with the Society for Science. The 2023 event featured more than 1,600 young scientists representing 64 countries and 49 U.S. states. Nearly $9 million was awarded in the form of cash awards and scholarships. Sigma Xi continued its support of these students as a Special Awards Organization during the 2023 event.

A group of Sigma Xi volunteers judged team projects in life sciences and physical sciences. The 11 judges evaluated 240 team projects to award the best demonstrations of team science and those reflecting Sigma Xi's culture of bringing together researchers from different disciplines. Congratulations to the following 2023 award winners:

Life Science

First Place — $1,500
O-RA: Osteoarthritis Rehabilitation Assistant
Sroy Kabkaew, Prince Royal's College, Chiang Mai, Thailand
Krittapat Tragoolpua, Prince Royal's College, Chiang Mai, Thailand
Napaschol Inthapan, Prince Royal's College, Chiang Mai, Thailand

Second Place — $1,000
DIVA: Spatial Navigation for the Visually Impaired Using Depth-Sensing Artificial Intelligence and Convolutional Neural Networks
Karthik Muthukkumar, Urbana High School, Ijamsville, Maryland, United States of America
Pranav Sristy, Urbana High School, Ijamsville, Maryland, United States of America
Tarun Malarvasan, Urbana High School, Ijamsville, Maryland, United States of America

Third Place — $500
Antimicrobial Activity of Organic Herbal Extracts Infused with Ozone as a Combination Therapy on Helicobacter pylori
Aalaa Yousuf Al Aufi, Al Amal School, Barka, Oman
Basma Khalid Al Uwaisi, Al Amal School, Barka, Oman

Physical Science

First Place — $1,500
Developing an LSTM-Based Model for Accurate Sign Language Recognition through Hand Keypoint Detection
Alex Antonio Hernandez Juarez, West High School, Salt Lake City, Utah, United States of America
Cole Nen Li Chu, West High School, Salt Lake City, Utah, United States of America

Second Place — $1,000
Anti-Forma Chitogel from the Hydrogel of Chitosan of Shells of Hermetia illucens
Guojun Wu, Carmel Pak U Secondary School, Hong Kong
Wai Ki Wong, Carmel Pak U Secondary School, Hong Kong

Third Place — $500
Intelligent Hydrogel Synthesis and Bracelet Design to Detect Foreign Materials in the Human Body and in Beverages
Ekin Asyali, Gaziantep Private Sanko College, Gaziantep, Turkey
Sude Naz Gulsen, Gaziantep Private Sanko College, Gaziantep, Turkey

Register Today!

IFoRE '23

POWERED BY SIGMA XI

INTERNATIONAL FORUM ON RESEARCH EXCELLENCE

NOVEMBER 10–12, 2023 | LONG BEACH, CA

- Student Research Competition and Awards
- Symposia
- Panel Discussions
- Workshops
- Networking Events
- Career Mentoring
- STEM Art & Film Competition and Awards

Featured Speakers

NOBEL LAUREATE



FRANCES ARNOLD
California Institute of Technology



SUZANNE SCARLATA
Worcester Polytechnic Institute



MIGUEL JOSÉ YACAMÁN
Northern Arizona University



JAMES M. SULLIVAN
Florida Atlantic University



RICHARD BOUDREAULT
McGill University

Sigma Xi's Annual Meeting and Student Research Conference is now IFoRE

experienceIFoRE.org

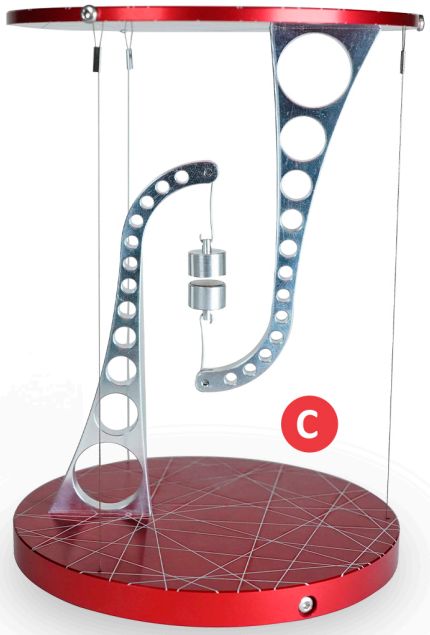
SCIENTIFICS DIRECT®

FREE SHIPPING ON ORDERS $79+ USE CODE: SDISSUESST823

FOR THINKERS WHO WANT TO LEARN

HOW AND WHY







A: #3157852 | Super Precision Gyroscope with Gimbal Kit
B: #3157495 | Galileo Watchmaking Kit
C: #3157863 | Magnetic Tensegrity Engineering Kit

Discover our Unbeatable Selection at www.ScientificsOnline.com